2013

Signal Processing of Electroencephalogram for the Detection of Attentiveness towards Short Training Videos

Paul Nussbaum
Virginia Commonwealth University

Signal Processing of Electroencephalogram for the Detection of Attentiveness towards Short Training Videos

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

By

Paul Alton Nussbaum
B.S., University of Rhode Island, 1987
M.S., NYU Polytechnic, 1994

Director: Rosalyn Hargraves, Ph.D., Department of Electrical And Computer Engineering

Virginia Commonwealth University
Richmond, Virginia
September 2013

Acknowledgements

My work on this research and dissertation was made possible with the support of a large number of people, both inside and outside of Virginia Commonwealth University (VCU). Within the faculty ranks I would like to thank my advisor Dr. Rosalyn Hargraves. Through her expertise and patient guidance, she taught me how to pursue research with a strong electrical engineering and biometric device foundation, while also touching upon my love of teaching. Dr. Kayvan Najarian guided my research and signal analysis methods, including the suggestion that I find a baseline method. With his deep knowledge of biomedical device systems and methods, he was also instrumental in my selection of VCU. I would also like to thank my dissertation advisory committee who has taught me a great deal. Private meetings with Dr. Aphroditi Filippas were never rushed despite her busy schedule. She took the time to explain the importance of thoroughness and accuracy of wording. Dr. Alen Docef was also a wonderful teacher, explaining simply and clearly how hardware implementation aspects should be measured and described. We are all sad to have lost Dr. David Primeaux who passed away suddenly, may he rest in peace, and we will all miss his insightful tutelage. Last on the list of faculty I would like to thank, and definitely not least, Dr. Paul Gerber joined the committee late and yet provided huge amounts of guidance in the format and structure of the educational research videos and questionnaires.

Fellow students like Ashwin Belle and Alfred Herrera and others have cheerfully provided me with guidance and helped me navigate the waters of EEG data collection. Fellow students were also supportive during the many steps along the way towards my Ph.D., and so were the wonderful staff at VCU, from Dean Boyan to Leena Joseph, Aisha Shabazz and many others.

In addition I would like to thank my dad William, may he rest in peace, and my mom Mary for their wonderful support, and to my daughter Nicole with her understanding of her dad's endless "homework." I want to thank my son Shawn, with his proof reading, psychology expertise and patience. More than any other person; I want to thank my wonderful wife Debbie. Without her managing the disruptions to the family and household caused by my keeping a full time job while enrolled in the rigorous VCU Ph.D. program, none of this research, or my happiness in life, would be possible.

Beyond all people I thank the Lord and pray for the wisdom to continue in good works.

Table of Contents

List of Figures

List of Tables

List of Equations

Abstract

This research has developed a novel method which uses an easy to deploy single dry electrode wireless electroencephalogram (EEG) collection device as an input to an automated system that measures indicators of a participant's attentiveness while they are watching a short training video. The results are promising, including 85% or better accuracy in identifying whether a participant is watching a segment of video from a boring scene or lecture, versus a segment of video from an attentiveness inducing active lesson or memory quiz. In addition, the final system produces an ensemble average of attentiveness across many participants, pinpointing areas in the training videos that induce peak attentiveness. Qualitative analysis of the results of this research is also very promising. The system produces attentiveness graphs for individual participants and these triangulate well with the thoughts and feelings those participants had during different parts of the videos, as described in their own words.

As distance learning and computer based training become more popular, it is of great interest to measure if students are attentive to recorded lessons and short training videos. This research was motivated by this interest, as well as recent advances in electronic and computer engineering's use of biometric signal analysis for the detection of affective (emotional) response. Signal processing of EEG has proven useful in measuring alertness, emotional state, and even towards very specific applications such as whether or not participants will recall television commercials days after they have seen them. This research extended these advances by creating an automated system which measures attentiveness towards short training videos.

The bulk of the research was focused on electrical and computer engineering, specifically the optimization of signal processing algorithms for this particular application. A review of

existing methods of EEG signal processing and feature extraction methods shows that there is a common subdivision of the steps that are used in different EEG applications. These steps include hardware sensing filtering and digitizing, noise removal, chopping the continuous EEG data into windows for processing, normalization, transformation to extract frequency or scale information, treatment of phase or shift information, and additional post-transformation noise reduction techniques. A large degree of variation exists in most of these steps within the currently documented state of the art. This research connected these varied methods into a single holistic model that allows for comparison and selection of optimal algorithms for this application.

The research described herein provided for such a structured and orderly comparison of individual signal analysis and feature extraction methods. This study created a concise algorithmic approach in examining all the aforementioned steps. In doing so, the study provided the framework for a systematic approach which followed a rigorous participant cross validation so that options could be tested, compared and optimized. Novel signal analysis methods were also developed, using new techniques to choose parameters, which greatly improved performance.

The research also utilizes machine learning to automatically categorize extracted features into measures of attentiveness. The research improved existing machine learning with novel methods, including a method of using per-participant baselines with kNN machine learning. This provided an optimal solution to extend current EEG signal analysis methods that were used in other applications, and refined them for use in the measurement of attentiveness towards short training videos.

These algorithms are proven to be best via selection of optimal signal analysis and optimal machine learning steps identified through both n-fold and participant cross validation.

The creation of this new system which uses signal processing of EEG for the detection of attentiveness towards short training videos has created a significant advance in the field of attentiveness measuring towards short training videos.

Novelty and Contribution

The research has led to a number of novel signal analysis and machine learning methods which are described in detail in the Methods chapter of this dissertation. The contribution of these methods to the final system, both quantitative and qualitative, is described in detail in the Results chapter of this dissertation. A brief summary of these novel methods are described here.

1. Creation of a signal analysis method specifically for this application

This is as a result of collecting data from a comparatively large number of participants and then using participant cross validation to compare a number of signal analysis parameters as well as noise removal processing algorithms. This resulted in the creation of a system and method for the use of EEG to detect attentiveness towards short training videos achieving an 85% and better rate of accuracy.

2. Ensemble Averaging of Attentiveness Indicators across Participants

The system created in this research calculates a numerical attentiveness indicator based on EEG data, and can graph that attentiveness information as it changes over time. On a per-participant basis, this can be triangulated with qualitative data such as participant self-reporting of what they were thinking and feeling at different parts of the video, or a specific time when a cellphone vibrated in the middle of a video. The research goes beyond this to also produce an ensemble average attentiveness graph, using data from all the participants who watched the same video. This is ensemble average can show the operator which specific parts of the video are generating greater attentiveness than other parts. In one video, for example, the ensemble average attentiveness goes up at a specific point in the video where the camera switches from a shot of the instructor's face, to a shot of the instructor's hands as he shows how the piece of paper is to be folded. This hands-on demonstration, or modeling, is well known in education to be important in student learning, and this research has provided biometric analysis that shows this with great precision. This novelty shows great promise for the use of this in future educational research.

3. Noise interval method of selection of thresholds for optimal muscle and ocular noise removal

Certain algorithms for muscular and ocular noise removal from EEG signals require threshold values to be selected. The threshold is used to identify spikes of voltage to be removed, and the current methods for selecting levels for these thresholds are not numerically based on the desired end result, that of removing infrequently occurring noise events. The proposed method is based on the expected average interval between these infrequent events by visually displaying amplitude levels of spikes that appear on average once per second, once per half second, etc… The selection of the average rate of noise occurrence is left to the operator, based on the fact that the operator is purposely trying to remove relatively infrequently occurring noise. Nevertheless, the novel method produces consistent results across a range of selected rates, and also is able to solve other issues of calibration over many experiments, as well as solve issues related to DC bias (differences between positive and negative threshold values).

4. Baseline Setting

Machine learning algorithms require training, or at least a fixed exemplary set of data points from which distance or linear transformation algorithms are based. Once the entire system has been developed, no further training is possible, because the system will be used by an operator who wishes to determine if a participant is attentive towards a short training video. The operator does not perform manual classification of the participant's EEG data, and therefore needs the automated system to provide this information. Nevertheless, as with other biometric data (such as taking your blood pressure consistently at the same time of day), it is possible for the operator to set a baseline for a particular participant. For example, if the participant is asked to sit still and look at a boring video, then the automated system can use this measurement as a baseline and scale future output of the machine learning algorithm up or down based on that single set baseline. Similarly, a double baseline can be created by having each new participant watch two different videos prior to use of the system. Whether single baseline or double baseline is used, there is no additional training of the system, and no special knowledge by the operator is needed. The novelty described in this research is the specific way baseline EEG data is used within the optimal machine learning algorithm.

5. New signal analysis transformation technique using gradient features (GF) instead of traditional Fourier or wavelet transform:

Most of the analysis in this research is performed using traditional frequency or wavelet methods, but a third experimental method was desirable, both to see if improvements could be made and to see if different methods agreed with each other with regard to participant attentiveness. This third experimental method called Gradient Features (GF). Motivated by control theory– signals and systems have been modeled using difference equations, or gradients. GF seeks to expand upon this and identify features of gradients, just as Discrete Wavelet Transform (DWT) identifies features of approximation coefficients. Before DWT data can be used as input to a classification algorithm, a series of coefficients must be converted into a handful of descriptive features. These may include Power, Mean, Median, Max, Min, Slope, Entropy, and others. In addition, DWT methods frequently use pre-filtering to divide the signal into bands prior to wavelet transformation. GF removes the need for DWT transformations, and the pre-filtering, and instead finds these same descriptive features about the raw amplitude over time data. GF then repeats this process over successive partial derivatives of that same raw data. This lends itself to features which are also computationally simple to extract and making it simpler to deploy on an embedded processing system – which may not have the multiplication speeds of a desktop CPU. Although modern CPU's calculate multiplications as quickly as addition and subtraction, this is not the case for simpler and less expensive embedded processors – which may benefit from such an algorithm, and can keep costs down if the solution is to be deployed to many students. Future research is suggested towards the optimization of GF to the development of this and other biometric devices.

6. Pseudo-Audio to hear EEG simultaneously with Short Training Video

By overlaying an audio signal that indicated attentiveness on top of the training video, an operator would find it easier to spot which portions of the short training video elicit the most attentiveness. Another benefit of pseudo-audio is that the audio playback can include actual EEG data, so the operator can listen to the EEG, or even the extracted features of the EEG. Automated classification systems for biometric data are complex and highly dependent on the signal analysis and feature extraction methods presented to the machine learning algorithms. These methods often seek to remove noise, but may also remove other important information. It is often not possible to visualize on a two dimensional display the details of these features with enough clarity for an operator to determine if they contain sufficient information for the machine learning algorithm to classify correctly. It is possible to perform inverse transforms on the feature data and thereby re-create a semblance of the original signal, which can be used by the operator to determine if sufficient information remains to manually categorize the signal. The novelty arises from the idea that it may be simpler for an untrained operator to use audio feedback, rather than visual, to hear the differences between different categories of EEG signals. Since EEG has several bands that are below the threshold of hearing, it is possible to speed up the EEG signal so operator can hear it. This can be superimposed on the short training video, which also plays faster, but can be easily followed by the user. The user can then clearly hear when the EEG changes due to specific events (such as modeling, where the teacher demonstrates a hands-on exercise). The method is simple and powerful, and might possibly be extensible to other biometric applications where visualization does not provide sufficient access and comprehension of the data.

7. Method of Selecting Mother Wavelets and Decomposition Levels

A rule of thumb in selecting the correct mother wavelet for a particular feature extraction exercise is to choose a mother wavelet whose shape is closest to the signal desired to be identified. The wavelet transform is essentially a cross correlation of the mother wavelet (at different scales) and the signal in question, so a good measure when comparing mother wavelet shapes would be a comparison of maximal values across different shifts and scales. This method is essentially a brute force method, simply trying a large number of combinations of mother wavelets and scales (decomposition levels) on the raw data and seeing which one has the highest maximum value across all time shifts. Critical to this novel method is the use of the same pre-processing algorithms as will be used in the automated system in question. The values are compared, and then the largest is selected to use in comparing wavelets, and decomposition levels for various EEG bands. The method is then validated against the complete system to see that it does indeed provide more accurate results than other wavelet parameters.

CHAPTER 1 Introduction

The dissertation is organized to describe the research and newly developed system and method which use electroencephalograms (EEG) as an input to an automated system that measures indicators of a participant's attentiveness while they are watching a short training video. Important due to the increased use of distance and self-learning methods, the research takes advantage of recent advances in EEG hardware and methods, and improves upon them to provide a valuable engineering solution to a timely educational need. The bulk of the research is focused on signal analysis of EEG for this application, although other areas are equally important to describe this complex problem. The document first provides an extensive background of the state of the art. This is then followed by a detailed methodology description and results analysis, as well as conclusions and suggestions for further investigation.

The background section of this document begins with a description of related research in the areas of training videos and the definition of attentiveness used in this research. The section then goes on to describe the many applications of EEG towards the classification of mental state currently being investigated by researchers around the world. Most importantly, the background section of the document compares in great detail the existing signal analysis and feature extraction methods used on EEG, and also the methods and number of participants used in related research. Finally, the background section explains the validation methods currently used in EEG mental state research. Throughout the section, tabular comparisons illustrate the variety of systems and methods employed today, including the different EEG sensor positioning, noise decontamination, windowing, frequency banding, number of participants, and accuracy validation algorithms.

The methods section of this dissertation described the experiment itself, as well as some of the subject selection details, since the use of human participants is required. Once the experimental procedure is described, the signal analysis methods are detailed, including a variety of novel methods. Signal analysis descriptions include noise decontamination, windowing, and transformation details for the solution. Machine learning and validation methods are also detailed.

Results are then presented along with analysis comparing quantitative and qualitative data. Also described are ensemble averages of attentiveness measures across all participants, and individual participant attentiveness graphs, which are also qualitatively triangulated with questionnaire responses and observation notes.

The results demonstrate that the developed system is suitable for use in real world situations outside of the laboratory. Conclusions and future areas of investigation are also provided.

CHAPTER 2 Background and Literature Review

2.1 Purpose and Overview

The motivation of for this research rests on both the ongoing need to improve educational multimedia content for distance learning and computer based training, as well as the opportunities presented by engineering advances in EEG signal analysis. Recent advances in EEG signal analysis have demonstrated that this biometric signal, traditionally used to detect sleep patterns and seizures, can be extended to provide measures of affective or emotional state. This research builds upon those signal analysis methods to extend these for practical use in measurement of attentiveness towards short training videos. Hardware advances in EEG have also made the physical devices lightweight, wireless, and easy to attach to the scalp without sacrificing application specific effectiveness. This research has used these recent advanced in EEG hardware and used them to gather extensive data. Finally, the research has developed several novel techniques and incorporated them into a system and method providing signal processing of EEG for the detection of attentiveness towards short training videos. To understand the research and its results, a background understanding of the prior state of the art is needed. This chapter seeks to provide this background information and describe the current state of the art prior to this research.

The area of study covers a wide number of research areas, and as such, this review of the existing body of literature requires a system of organization that will allow for an understanding of how the topics fit together, and where prior studies set the foundation for this research. A diagram outlining the organization of this review of related work is shown in Figure 1. Using this structure of organization existing research can be compared and contrasted, and also interdisciplinary research can be brought together in an orderly manner for the purposes of describing the background of this research.

These inter-disciplinary topics include:

- Attentiveness towards short training videos
- EEG for the classification of mental state
- Signal processing of EEG data
- Pattern Matching / Machine Learning / Classification algorithms
- Validation techniques used to gauge effectiveness of methods and algorithms

A review of self-paced asynchronously delivered training via video is discussed. Existing research on the use of EEG biometric measurements for the detection of attentiveness and other similar applications is also examined. Furthermore, a variety of described techniques are examined in the areas of pre-processing related to EEG analysis, for the specific purpose of feature extraction. Pattern matching algorithms are needed for the machine learning to classify the extracted features into categories of attentiveness and non-attentiveness. Finally, a variety of techniques that have been employed to validate methods and algorithms for automated EEG analysis are examined. Therefore, this background chapter examines all of these disciplines of research, as shown in Figure 1.

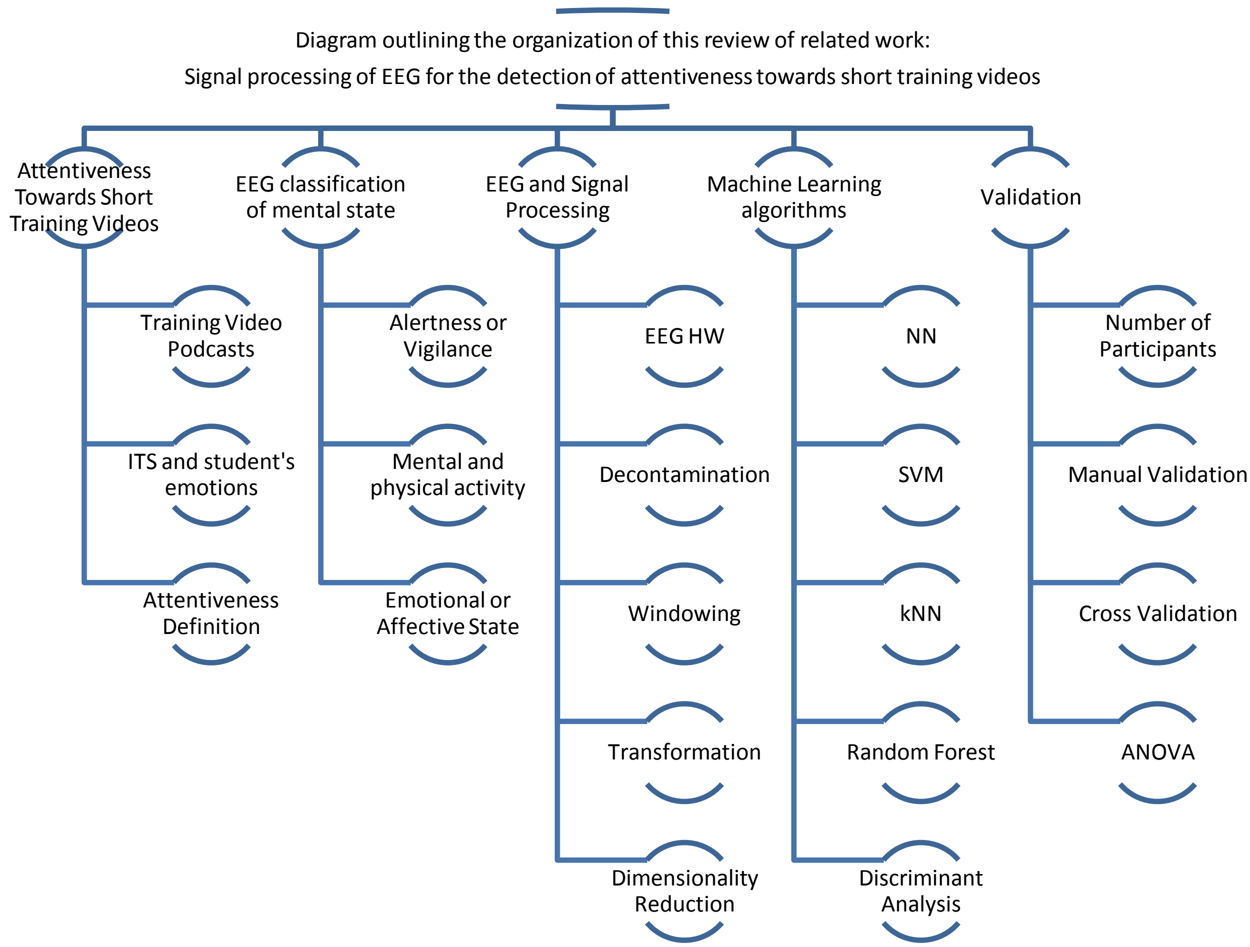


Figure 1 Organization of Literature Review

2.2 Training Videos

This section provides a brief review of training videos. An overview of recent research into the use of training video podcasts by students is provided. The section then concludes by covering Intelligent Tutoring Systems (ITS) which use videos and attempt to engage students into a more active learning method, even to the point of trying to detect the student’s emotion and adjust the training accordingly.

2.2.1 Training Video Podcasts

As distance learning and computer based training become more popular, published research in this area has increased. Researchers (Bakera, et al. 2010) explain that many factors can impact how well computer based training is received. The format of the lesson can greatly change how actively the student is engaged, as well as many other factors including culture, age, personality, and gender factors, and also the context (e.g., day before an important exam) and subject preference (such as if one student likes history more than algebra).

The delivery of computer based training through the use of videos is of great interest. Considerable investigation and experimentation has been conducted on video training. (Kay 2012) presents an extensive overview of current research into training videos which are sometimes called podcasts (after a popular publication method of the same name). It was shown that students choose to use podcasts because of improvement and control of their learning, as well as catching up on missed lectures. Research also indicates that these video podcasts are emotionally enjoyable to the students, as well as being satisfying and motivating. (Kay 2012) describes video podcast research as finding largely positive cognitive results, and unique benefits such as convenience (e.g., choosing a time, location, access method), and ability to access lectures from experts geographically far away. (Kay 2012) notes the positive impacts of video podcasts on student behavior, including frequency of viewing and improvement of study habits such as more independence, more self-reflection, more efficient test preparation, more reviewing of material, and increased contact with academic staff. The research also shows that enhanced podcasts like audio narrated slideshows led to improved multiple choice test results. Also, the research demonstrated a positive correlation between the uses of worked example video podcasts, which are clips designed to explain, articulate and assist students in learning how to solve specific problems, and test scores (Kay 2012) which further demonstrated that video training and podcasts have a positive impact on learning.

(Freire, Lopes and Campos 2012) examined the complex ways in which e-learning systems have evolved. Their paper discussed learning management systems, and not short training videos specifically, but the discussed systems are meant to store, manage, and modify such content as well as many other educational content materials. The authors concluded that usability is an area of continuing importance and research, including aspects such as satisfaction, user motivation, engagement, and interaction (involvement). In addition, the authors concluded that there is a need to improve usability measurement tools, especially with regard to making them quicker and more complete.

It is therefore evident that a timely and significant need exists for a system that can measure the quality of educational materials, such as short training videos, and that such a method should measure engagement, interest, and involvement.

2.2.2 ITS and Student's Emotions

Besides simply measuring engagement, interest, and involvement; there is also a need to gauge the affective emotional state of a student.

Research on the discussion of Intelligent Tutoring Systems (ITS) (Bakera, et al. 2010) suggests that the Human-Computer Interface (HCI) of a Computer Based Training (CBT) system should modify its behavior dynamically based on the current emotional state of the user. This change in how the ITS behaves is in addition to the more traditional ITS dynamic modifications, i.e., those based on identification and correction of student errors and other assessments. The paper also described the considerable research examining the affective states of a user (i.e., emotions, moods, feelings) and the importance of measuring this to develop more effective, user-friendly applications. The authors go on to explain that this is due to a complex interplay between cognition and emotion. When it comes to duration and impact of the affective state of the user (Bakera, et al. 2010) the researchers conclude that boredom is one of the most persistent emotions and therefore boredom is important to avoid since it is difficult to overcome, and yields poorer learning results. This is because frustration and confusion, although not seemingly “positive” emotional states, are quickly changed, and may not need ITS remediation, because these emotional states can still end up yielding good learning results.

To conclude research has shown that video lessons and podcasts have been shown to have a positive impact on learning as demonstrated through prior research. Also, in the case of the video lesson, there exists a pressing need for systems that measure the engagement and emotional state of the user which may be critical to the efficacy of the training. It is therefore of interest to determine the attentiveness of a student towards a training video.

2.2.3 Attentiveness definition

For the purposes of this research, the affective state of the participant watching the short training video is deemed to be attentive when the participant has a high positive affect, including feelings of satisfaction, engagement, interest, and involvement. This definition of educational attentiveness is not the only possible definition. The above definition is justified by related research in similar applications such as in the research cited above, as well as other related research in similar applications which are listed below in a brief chronological history, supporting the above definition of attentiveness.

In (Watson, Clark and Tellegen 1988) the authors propose a brief and easy to administer mood scale called the Positive and Negative Affect Schedule (PANAS), where the orthogonal axes of emotion are “negative affect” and “positive affect” and high positive affect is associated with terms such as high energy, full concentration, and pleasurable engagement.

In (Lester, Towns and Fitzgerald 1999) affect is explored not from the point of view of the learner having an emotional response, but rather a computerized tutoring system having an avatar that uses facial expressions and body movements to communicate an affective state to the learner. With regard to the range of emotions expressed by the avatar, the researchers explained “…because their role is primarily to facilitate positive learning experiences, only a critical subset of the full range of emotive expression is useful for pedagogical agents. For example, they should be able to exhibit body language that expresses joy and excitement when learners do well, inquisitiveness for uncertain situations (such as when rhetorical questions are posed), and disappointment when problem-solving progress is less than optimal.”

In (Kort, Reilly and Picard 2002) the authors defines a model of emotions and learning that explains the emotional state of the learner as being in one of four quadrants, depending on the positive or negative value on two axes. The first axis is “learning,” and the second axis is “affect.” The positive side of the affect axis is associated with terms such as awe, satisfaction, curiosity, and hopefulness.

In (Chalfoun, Chaffar and Frasson 2006) the researchers use machine learning to predict a participant’s self-assessed emotion when presented with the results of a quiz and personality test. The participant finds out the results, and then expresses their emotions by selecting one of the following words: disappointment, distress, joy, relief, satisfaction or fear-confirmed.

In (D'Mello, Craig and Graesser 2009) researchers analyze the affective state of participants while they are performing a learning task with a computerized natural language tutor using both online and offline self-reports by participants as well as peers and trained judges observing facial features. Words that the study used in both self-reports and reports by observers (and their definitions) include those shown in Table 1.

Affective state	Definition
Boredom	State of being weary and restless through lack of interest
Confusion	Failure to differentiate similar or related ideas/ noticeable lack of understanding
Flow	State of interest that results from involvement in an activity
Frustration	Making vain or ineffectual efforts however vigorous; a deep chronic sense or state of insecurity and dissatisfaction arising from unresolved problems or unfulfilled needs; dissatisfaction or annoyance
Neutral	No apparent emotion or feeling
Surprise	Wonder or amazement, especially from the unexpected

Table 1 Affective state of participants as they use an ITS

It is useful to note that the aforementioned study found the Kappa of trained judges who watch the facial expressions of the participants and code facial actions to provide affective judgments was only 0.36.

Although Kappa is a conservative measure of agreement, it still shows that the affective states of participants are difficult to judge even by those who are trained to do so. The Mathematical definition of Kappa is shown in Equation 1.

$$\boldsymbol{\kappa = \frac{\mathrm{Pr}(a) - \mathrm{Pr}(e)}{1 - \mathrm{Pr}(e)}}$$

Equation 1 Kappa measure of agreement

Where $\mathrm{Pr}(a)$ is the relative observed agreement among raters, and $\mathrm{Pr}(e)$ is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category. If the raters are in complete agreement then $\kappa = 1$. If there is no agreement among the raters other than what would be expected by chance, as defined by $\mathrm{Pr}(e)$, then $\kappa = 0$.

In a later study on the same topic (D'Mello and Graesser 2012) the issue of disagreement between trained judges is removed by having participants self-rate their affective state. In (D'Mello and Graesser 2012), the definitions of affective state needed to be explained to the untrained judges (the participants themselves) and the definitions were explained to the participants using wording similar to before, as shown in Table 1.

In (Heller, et al. 2010) the researchers attempt to dive deeper into the affective state of the learner by gathering, in their own words, the definition of what they mean by “engagement in courses” and “What makes a course engaging to you?” The learners in this study were male and female freshman and sophomores in engineering, and the researchers were able to group the responses into categories, as well as give specific wording terms that the students used. Some notable popular categories and terms used include: active participation, hands on, faculty enthusiasm and interest, discussion, interaction between faculty and students, as well as other similar terms, and other less popular terms.

The aforementioned related research therefore supports the use of the definition for this research of attentiveness towards short training videos as having a high positive affect including feelings of satisfaction, engagement, interest, and involvement.

2.3 The use of EEG Data for the Classification of Mental State

The EEG is one of the most useful tools in clinical neurophysiology. EEGs are voltage measurements of the scalp, representing the sum of synchronous postsynaptic potentials arising from broad cerebral cortical areas and can be used for the identification of cerebral injuries or disorders (Epstein 2012). Research also shows that EEG data can be used to recognize other more subtle mental states. Although a very wide variety of applications are described, the literature does not involve signal analysis of EEG data being specifically used to measure attentiveness to short training videos. Nevertheless, the wide cross sections of published applications that have been researched have laid the foundation for the current research to be successful through refinement of the signal processing and pattern recognition techniques.

2.3.1 Alertness or Vigilance

Alertness and vigilance mental states are well studied with regard to EEG data correlation. The published research of (MacLean, Arnell and Cote 2012) shows how EEG data from participants who are resting can later be used to predict how well they can perform during fast paced target identification using “attentional blink” measures. (Goldfine, et al. 2011) demonstrated that EEG analysis can reveal awareness in brain injured patients who are otherwise unable to communicate, but who are asked to mentally imagine motor and spatial navigation tasks. There has been research on using EEG to detect when someone is no longer alert enough to safely operate a vehicle or maintain display vigilance (Wilson and Bracewell 2002). Another example (Jung, et al. 1997) used EEG data to predict alertness as measured by lapses in auditory and visual sonar detection by trained Navy participants. Human experts can also look at features extracted from EEG data and tell if the participant is alert versus asleep or drowsy, as in the case of (Subasi, et al. 2005) where trained neurologists looked at the EEG recordings, and then picked which EEG sequences clearly indicated alert, drowsy, or sleepy states of the subject.

2.3.2 Mental and Physical Activity

More detailed classification of mental activity through EEG has been researched (Khare, et al. 2009) including differentiation between relaxed, imagining moving the right hand, watching a figure being rotated (imagining it as well), trivial multiplication (i.e., 2x3), and non-trivial multiplication (i.e., 49x78). Other mental tasks can be classified using EEG data (Palaniappan 2006) to determine the mental activity of participants from a series of mental tasks including geometric figure rotation, multiplication, writing a letter to a friend, visual counting, and resting.

2.3.3 Emotional or Affective State

EEG data has been used to classify the emotion of the user (Chakraborty, et al. 2009) as compared to their facial expression when watching emotion inducing movies, where 50 participants were shown 60 audio-visual clips, covering a range of six different emotions (anxiety, disgust (anger), fear, happiness, sadness, and relaxation). Similarly in (Khosrowabadi, et al. 2010) participants are presented with pictures and music to elicit four basic emotions of calm, happy, sad and fear, and they are also questioned using a self-assessment manikin as to the emotion they felt. The emotion is mapped onto a Valence-Arousal emotion plane and EEG data is then measured to see if emotion can be detected. In (Murugappan 2011) EEG data are related to emotional state by using five videos intended to elicit emotions (disgust, happy, fear, surprise and neutral) as selected by a college student panel from 115 international standard emotional clips. Separate participants wear EEG devices and the data is used to identify the emotion elicited by the video clips. In a more recent study, (Uusberg, et al. 2013) examined how certain frequencies like Alpha band (discussed below) of EEG activity may be more related to affective emotional state than earlier thought. (Uusberg, et al. 2013) shows that the alpha band is present in varying amounts depending on the affective stimulus (pictures from the International Affective Picture System). It was found that the prevalence of details in the picture affected the Alpha band, and as such cannot be discounted when comparing affective stimulation results. Also it was found that aversive (powerfully unpleasant) images generated high Alpha power, compared to rest and other images.

Some studies try to use EEG data to detect less dynamic mental states such as personality, or pathologies such as mental disorders. In one study (S. Lee, B. Abibullaev and W.-S. Kang, et al. 2010), children between the age of 10 and 13 years with Attention Deficit Hyperactivity Disorder (ADHD) were assessed for mental retardation using the Korean

Educational Development Institute's Wechsler Intelligence Scale for Children. Afterwards EEG data was collected. Participants were asked to relax, count, and perform mental tasks. The EEG data was used to recognize if the child had only ADHD, or also had mental retardation. One research paper (Ito, et al. 2010) demonstrated the ability to determine the personality of participants by using EEG measurements. In the aforementioned study, EEG data was collected while participants listened to music segments which they later rated as "Matches Mood", "Does not match mood", and "Borderline" (Ito, et al. 2010). Separately, the personality score of the participant is known from a 50 question exam (rating the person in the 5 areas of "Critical parent", "Nurturing parent", "Adult", "Free Child", and "Adapted Child".) The authors then compared the false-detection accuracy of the EEG classification system of "mood matching" to the personality trait of the participant, and make the claim that that false-detection may be related to the person's personality (Ito, et al. 2010).

Even before the widespread use of EEG, there was an interest in the physiological and anatomical structure of the human brain with relation to purchasing decision making. Termed "Neuromarketing" in the seminal book of the same name (Renvoise and Morin 2007), the authors foreshadowed modern medical signal analysis techniques applied to market studies. Neuromarketing research includes such trials as (De Vico Fallani and al 2008) where EEG data is used to predict whether a short advertisement video clip embedded within a television documentary will be remembered or not several days after viewing.

EEG data has been used in some aspects of the education field. For example, (Rebolledo-Mendez, et al. 2009) researchers described using EEG data while a student is taking an examination. In this case, participants interact with a computerized avatar that asks multiple choice questions. EEG attention rating from 0 to 100 is correlated with speed/accuracy of response and user self-described attentiveness levels. There was no student education or training provided – only assessment - the questions were targeted at Computer Science, and participants all were majors in computer science (and so should have been familiar with the answers to the questions). In another research experiment (Crowley, et al. 2010) participants wearing an EEG cap were asked to do an increasingly stressful (faster) Stroop Test. A Stroop test is an intentionally cognitively demanding task that asks the participant to name the displayed color of a conflicting display word. For example the participant is presented with the word “red” but it is colored in a green font. Subsequently, people will unintentionally say “red” instead of the correct response which is “green.” During each trial the speed was increased, while EEG was used to register “dips in meditation" which suggests lower relaxedness. The same research paper describes using the Tower of Hanoi problem (stacking different sized disks) was given to participants three times in a row – the researchers explain that this problem appears difficult to solve but once a participant learns the stacking algorithm it becomes easier to solve (Crowley, et al. 2010). EEG data was able to show trends in lowered assessment stress. Another research example that correlates educational assessment exercise stress to EEG data is (Mostow, Chang and Nelson 2011) where easy and hard “reading for meaning” tasks assigned and EEG data compared for adults and children. Yet another research study (Berka, et al. 2004) correlates EEG data to mental states including alertness (simulation defending against incoming enemy planes), cognitive task (response rate in identifying the number "5" out of digits presented), and memory (recognizing memorized images).

Somewhat related to the current research, EEG data can also be used for the detection of physical activity (not just mental activity). In (Nagashino, et al. 2002) EEG is used to discriminate whether a participant in a relaxed state has their eyes open or closed, and in (Selvan and Srinivasan 1999) EEG data includes ocular artifacts [electrical signals from the muscles used to move the eyeball] where in this case the goal is the removal of those artifacts (adaptive noise canceller).

To conclude the review of related work on the use of EEG data to determine mental state, prior research describes a wide variety of applications for this electrical engineering biometric measurement. These include measurement of dynamic mental activity such as alertness, type of mental task, memory of a television commercial and emotional state, as well as more static mental aspects such as pathology [mental retardation, and ADHD] and even possibly personality. The review of related work did not find published research suggesting that signal analysis of EEG data has been used to measure attentiveness to short training videos. Nevertheless, the very wide variety of applications described in the literature have paved the way for this research to refine the signal analysis and pattern matching methods to extend the use of EEG to this new application.

2.4 Signal Analysis and Feature Extraction of EEG Biometric Data

This section is the major focus of this background chapter and the research as a whole. Signal analysis of EEG biometric data involves several steps which can be grouped into two major areas. The first area involves the collection of EEG data, and the second area of algorithmic procedures involves the conversion or transformation of EEG data into vectors that will later be presented to an automated classification mechanism. Figure 2 shows these two major areas diagrammed as two horizontal sequential processes, with the top row diagramming the sequence required for the collection of EEG data, and the bottom row diagramming the sequence of operations for transformation into vectors.

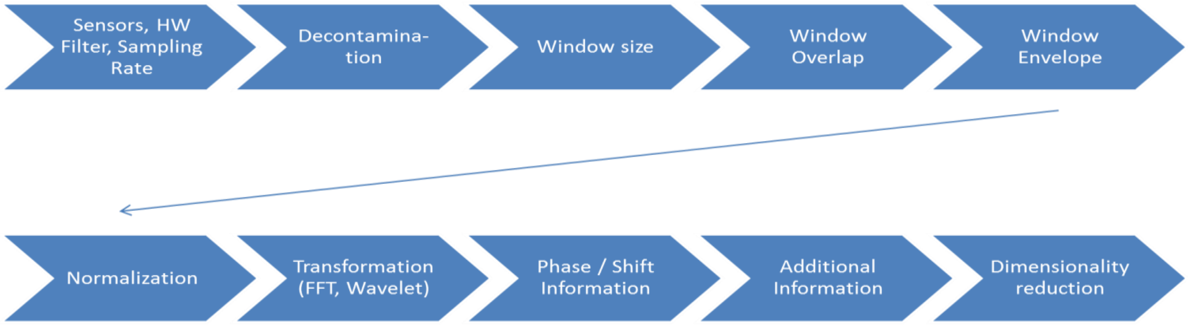


Figure 2 Detail of the components of the Signal Analysis of EEG Data

2.4.1 Sensors, Hardware Filtering, and Sampling Rate

EEG sensor positioning is often referred to according to a standard positioning chart (BIOPAC Systems 2011), often with the ground reference provided by the earlobe. Figure 3 shows a common EEG cap (left) as well as the International 10-20 system used to describe and apply EEG electrodes to the human head. Application of the sensors is time consuming. Notice the jar of electro gel and syringe applicator, the large number of electrodes that must each be gelled and applied, and the extensive amount of cabling (movement of which can cause unwanted noise to appear on the collected signals).

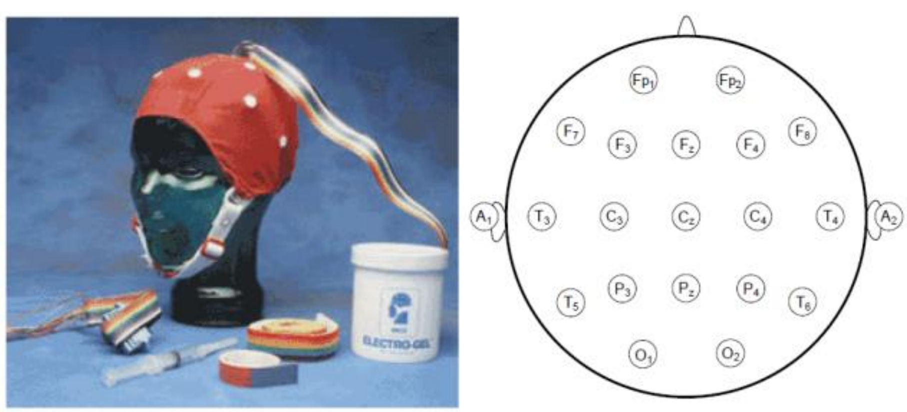


Figure 3 EEG cap and International 10-20 System (BIOPAC Systems 2011)

The selection of the sensor positions vary in the literature, although there seems to be a predominance of two types of sensor positions – frontal only or multiple sites on the scalp. The frontal positions are frequently though not exclusively used for time varying analysis. The all-over the head sensor positions involve a large number of sensor readings and are used if spatial analysis is being performed. The following sections provide a review of related experiments that document the sensor positions, pre-digitization filtering and sampling rate, and then the digital signal processing algorithms used.

Some examples in literature where time varying signals are of interest and electrode placement is described include (Chakraborty, et al. 2009) (Selvan and Srinivasan 1999) (Jung, et al. 1997) (Ito, et al. 2010) (S. Lee, B. Abibullaev and W.-S. Kang, et al. 2010) (Rebolledo-Mendez, et al. 2009) (Crowley, et al. 2010)and (Mostow, Chang and Nelson 2011). In (Chakraborty, et al. 2009) where researchers look to classify emotions, EEG data from frontal positions F3 and F4 are used. In (Selvan and Srinivasan 1999), where researchers are looking to capture and filter out ocular movement artifacts, use EEG electrodes F7 (frontal) and T3 (temporal, or at the temple of the head above the ear) based on the assumption that ocular noise is contained in this location. Additional electrodes are placed to collect electrooculography (e.g., eye movement) data to the left and right of eyes, and also above and below the right eye and the participant is asked to blink or rotate eyes while being monitored with EEG. In (Jung, et al. 1997) where alertness is measured, EEG data from two scalp sites ("two midline sites, one central (Cz) and the other midway between parietal and occipital sites (Pz/Oz), using 10-mm gold-plated electrodes referenced to the right earlobe.") are used. In (Ito, et al. 2010) where mood is being researched, a simple EEG cap (earlobe referential and frontal Fp1) is used to collect data. In (S. Lee, B. Abibullaev and W.-S. Kang, et al. 2010) where mental disorders are sought to be classified, EEG data was collected from 2 frontal lobe points.

Spatial analysis literature examples describing electrode placement include (De Vico Fallani and al 2008) and (Murugappan 2011). In (De Vico Fallani and al 2008) researchers examined a spatial network efficiency measure by examining 14 regions of interest (averaging current magnitudes of dipoles in each region) manually segmented from the cortical model of each subject (from MRI's). From this a spectral measure used to determine the directed influences between any given pair of signals in a multivariate dataset. In (Murugappan 2011) where emotional response was classified, participants wore a 64 electrode EEG cap and the data was analyzed on a spatial basis by first taking the Surface Laplacian of individual electrodes and their adjacent neighbors, where Surface Laplacian is defined in Equation 2.

$$\boldsymbol{X_{SLn}(t) = X_n(t) - \frac{1}{N_E}\sum_{i=1}^{N_E} X_i\,(t)}$$

Equation 2 Surface Laplacian

where X_{SLn} is the Surface Laplacian at a given electrode *n*, and N_E is the number of neighbor electrodes. Once this spatial analysis is complete, the remaining analysis described in (Murugappan 2011) is similar to other discrete time signals (and transformations of discrete time signals) described further below.

2.4.1.1 Justification of Use of Single Dry Electrode for This Research

Not all experiments are focused on collecting data from multiple locations on the head. Some EEG measurement devices seek to avoid a bulky EEG cap, liquid gel conductive material on the leads, and bundle of wires connecting the sensors to the digitization equipment. If the participant is more concerned about cold gel-like sensation in their hair, and keeping still in a chair because they are tethered to a big box, then they may not be focused on the experiment. For this reason, some models of EEG hardware have dry contacts (not requiring wet conductive gel) and wireless (not requiring cables tethering the user to the computer equipment) headsets. Figure

4 shows one such device from NeuroSky, the "MindWave Mobile" (NeuroSky Co. Feb 11, 2012). Using a wireless headset means that the bandwidth of data transmission greatly limits the number of sensors (as shown, there is only one forehead sensor measuring electrical potential compared to the earlobe ground reference clip). The dry sensor theoretically does not have the sensitivity of the wet sensor, although research described below shows this is not so much of a disadvantage when measuring attentiveness – perhaps in part because this is one of the few standard sensor positions that is not attenuated with hair, and has instead direct sensor to skin contact. It should also be noted that even though the dry contact single sensor headset is less obtrusive to the user, not requiring the application of cold wet gelatin, it is a device that must be placed on the head similar to wearing a set of headphones.

Use of a single frontal lobe EEG electrode is commonly described in literature. In (Frasson and Chalfoun 2010), the participant's affective state while using an Intelligent Tutoring System (ITS) is measured using a single frontal lobe EEG electrode with earlobe grounds showing an accurate detection (82%) of affective state from the brainwaves. (Frasson and Chalfoun 2010) was based on earlier work (Heraz and Frasson 2007) which also used the same EEG sensor setup and revealed an ability to use brainwaves to accurately predict participant self-assessed emotion according to the Self-Assessment Manikin (SAM) scale. (Belle, Hobson Hargraves and Najarian 2012) also uses only a frontal electrode (two electrodes are used, however the difference in voltage between the two is used as the only single EEG data stream over time) to achieve an 85.7 percent accuracy in identifying if the participant is watching an interesting video versus an uninteresting video. In (Chakraborty, et al. 2009), the single electrode is placed at F3 or F4 alternatively in an experiment that also collected facial expressions when watching emotion inducing videos, revealing that noisy correlated EEG-facial expression data could also be correctly recovered in as high as 95.2 percent of cases.

Recent research has set the precedent for use of a dry electrode in similar applications. Described is the use of a single electrode in the frontal position, this time without the use of wet conductive gel, specifically the MindSet EEG data collection headset made by manufacturer NeuroSky. In (Rebolledo-Mendez, et al. 2009), the same make and model of EEG data collection device as used in this experiment, showed a positive correlation between EEG data and participant self-reported attention during an experiment where the participant interacts with a computer generated three-dimensional avatar. Also using the same make and model of EEG data collection equipment is (Crowley, et al. 2010), where 78% of the time the brain wave data accurately identified the participant's self-categorization of category of affective state during an experiment which presented increasingly stressful mental challenges to the participant. (Mostow, Chang and Nelson 2011) also used the same device to predict with accuracy levels significantly better than chance whether or not the participant was reading an easy or a hard sentence. Studies also show the effectiveness of dry electrodes at various positions on the scalp (Gargiulo, et al. 2010) showing the signal recorded by a dry electrode is almost identical to that recorded by standard wet electrode EEG sensors.

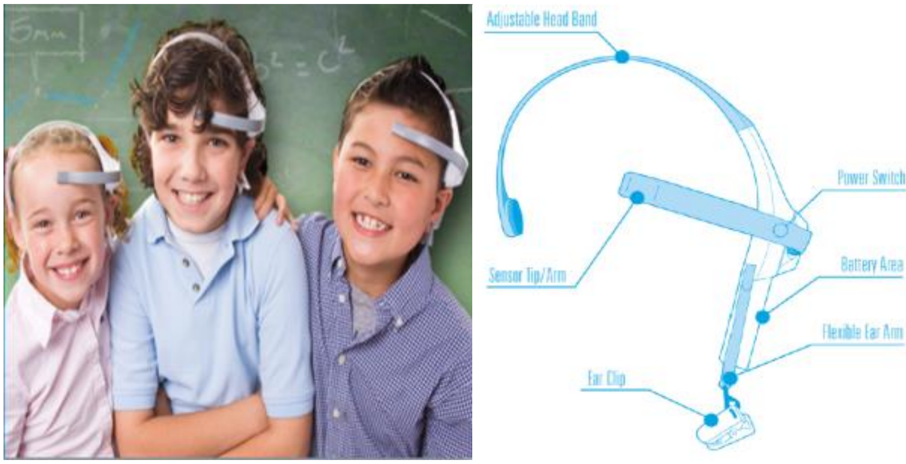


Figure 4 The MindWave Mobile from NeuroSky (Corporation Feb 11, 2012)

This research uses a single dry frontal electrode for the capture of EEG data, a reasonable EEG data collection method justified by the related research in similar applications.

2.4.1.2 Overview of different EEG sensor positions in Prior Research

Table 15 in Appendix A summarizes the different EEG sensor positioning methods selected by researchers.

2.4.1.3 Digitization Hardware

Besides sensor location, EEG data collection hardware also plays a role in the EEG signal that is examined. Digitization of the time varying data is accomplished using equipment such as the Biopac MP 150 (Figure 5) data acquisition system. With this equipment, the signal conditioning and filtering is already accomplished before digitization along with modifiable configuration settings (BIOPAC Systems 2011). The same functions are performed with the hardware used in this research (NeuroSky Co. Feb 11, 2012). Figure 6 shows the configuration settings available on the hardware used for this research, including output baud rate, power up data content, and notch filter setting (60 Hz for United States use products).

Figure 5 The Biopac MP 150 data acquisition system

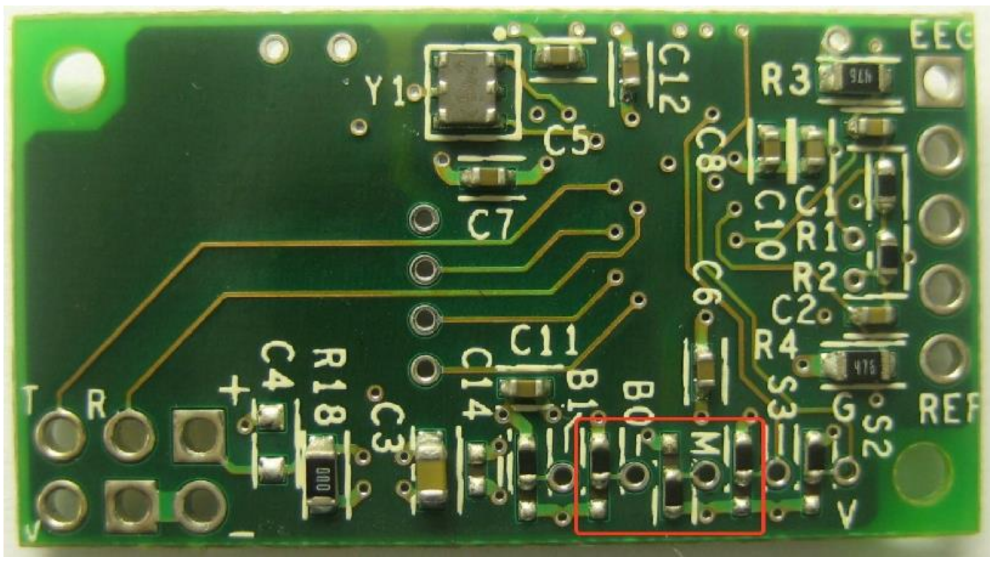


Figure 6 The TGAM1 board from NeuroSky showing the configuration pad outlined in red.

Before the data can be analyzed, it must be converted for storage on a digital computer. Some literature mentions the details of the parameters used for that conversion (pre-digitization filtering and sampling rate) and these parameters are briefly discussed below.

Filtering prior to digitization is performed to avoid sample aliasing by removing high frequency signals that would violate the Nyquist rate, which dictates that the sampling rate be twice the highest frequency contained in the EEG signal. Pre-digitization filtering is mentioned in (Subasi, et al. 2005) where band pass filters between 0.3 Hz and 70 Hz are used prior to digitization, and (Khosrowabadi, et al. 2010) mentions pre-filtering in the 2 Hz to 30 Hz range. The user manual for the Biopac hardware (BIOPAC Systems 2011) discussed optional settings including band pass between 0.1 Hz and 100 Hz, or band pass between 0.5 Hz and 35 Hz, with high pass and low pass filters available at those high and low frequencies respectively. Attentiveness is often associated with Beta wave and lower frequency activity, ranging up to 30 Hz (NeuroSky Co. Feb 11, 2012) and so the setting of pre-filtering in that range is supported.

Digitization sampling rate is mentioned in (Gwin, et al. 2010) at 512 Hz, (Nagashino, et al. 2002) at 200 Hz, (Khosrowabadi, et al. 2010) at 250 Hz, (S. Lee, B. Abibullaev and W.-S. Kang, et al. 2010)at 22.5 Hz, (Murugappan 2011) at 256 Hz, and (Mostow, Chang and Nelson 2011) at 512 Hz. The B-Alert software that comes with the Biopac hardware (BIOPAC Systems 2011) intended to detect the alertness of the participant, suggests a sampling rate of 256 Hz, which is consistent with the frequency of sampling needed to calculate band power accurately in the frequency ranges needed. Note also in the wavelet discussion below, there is a calculation simplification that arises by having a sampling rate a power of 2.

2.4.2 Decontamination of the Data

Traditionally studies have used extensive decontamination and removal of unwanted data from EEG including removal of ocular artifacts, power line signals, and spurious signals occurring from the movement of the sensors. These are signal contaminators in a normal EEG signal, and may require manual or automated removal. For example, the B-Alert system (Berka, et al. 2004) which samples EEG data at 256 Hz, first performs removal of artifacts caused by bumping or tapping of the sensors by searching for 3, 5, or 7 point data spikes with amplitudes greater than a threshold (40 mV in this case), as well as signals with amplifier saturation (maximum or minimum values of the digitizer) and excursions that occur before and after saturations. These are all removed by searching for the first zero crossings before and after these spurious signals, and replacing all data values between them with 0 μ V values. After this, a 60 Hz notch filter is used to remove power line signals (the experimental equipment and environment is exposed to these power current frequencies). Removal of ocular artifacts (blinking of the eyes) is more complex. In (Berka, et al. 2004) the EEG signal is passed through a 7 Hz low pass filter and then applying cross-correlation analysis to the filtered signal using the positive half of a 40 μ V 1.33 Hz sine wave as the target shape, and applying thresholds to the outputs of the cross-correlation analysis. Cross-correlation analysis is similar to convolution, except neither of the signals is reversed in time, as in Equation 3.

$$(E \star PS)[n] = \sum_{m=-\infty}^{\infty} E[m]PS[n+m]$$

Equation 3 Discrete Time Cross Correlation

E is the filtered EEG signal, and PS is the half-wave rectified sine wave. Once the start and endpoints of blink signals are identified (using minima and maxima analysis in each direction from the point of maximum cross-correlation) the blink signal is removed by replacing that segment of the EEG signal. Maxima and minima analysis is performed by looking for zero first derivatives, or in the digital sampled case, minimum difference between samples. Figure 7 and Figure 8 graphically depict a recreation of the process described in (Berka, et al. 2004) except using EEG data collected using the NeuroSky hardware (the single dry sensor on the forehead and earlobe ground samples at 512 Hz).

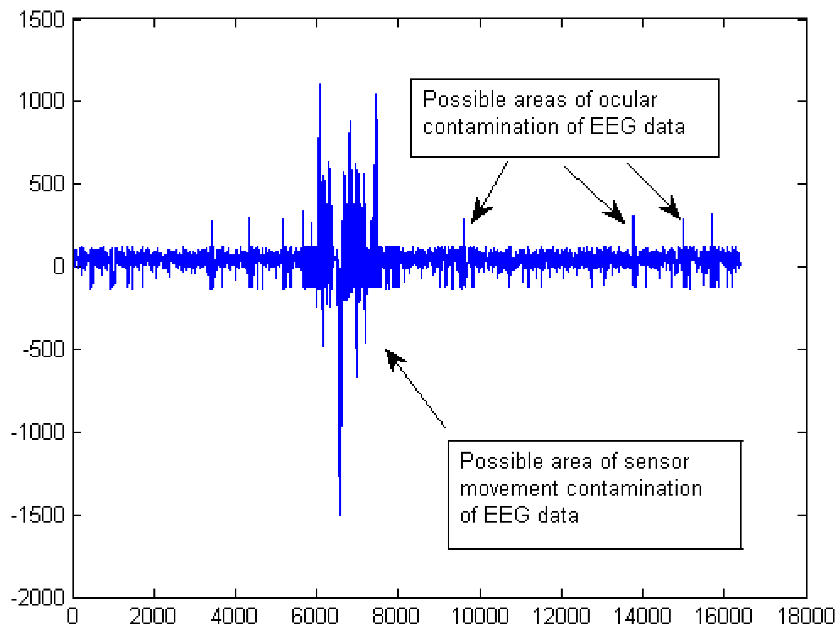


Figure 7 Raw EEG data sampled at 512 Hz showing possible areas of noise contamination

Figure 7 shows some of these possible areas targeted for decontamination. The sensor movement decontamination would be identified by the large amplitude of the signal, and replaced by zeroes. The removal of ocular artifacts requires cross-correlation and is graphically shown below in Figure 8, which shows, from left to right, a segment of the EEG with possible ocular artifacts, a one period of a 1.33 Hz half wave rectified sinusoid, and finally the cross-correlation of the two signals, showing identification of start and end of ocular artifacts (local maxima on either side of large minima). The "blink" segment consists of samples 3000 through 5000 from the previous EEG signal. The use of the positive portion of a cosine centered on 0 (instead of causal sine with only positive values of samples *n*) would have more precisely identified the blink ocular artifact.

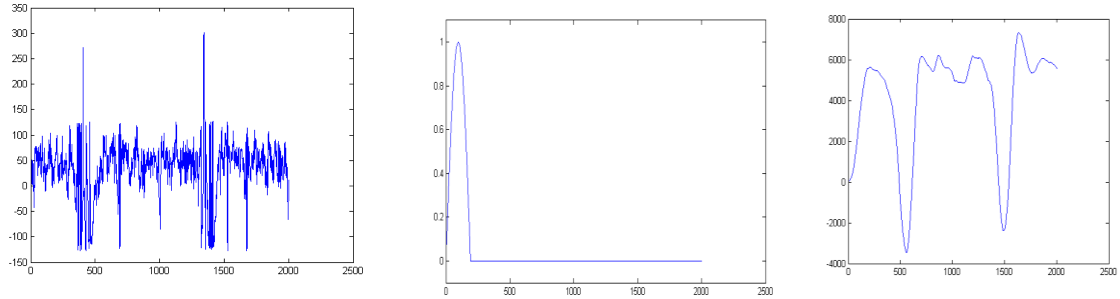


Figure 8 a) EEG signal with Ocular artifacts, b) one period of half-wave rectified sin(2π1.33) , and c) resulting signal after cross correlation.

Extensive as the above described decontamination algorithm is, it is by no means the only example in literature of EEG preprocessing. One simple method in (Jung, et al. 1997) used EEG signals sampled at 312.5 Hz and a threshold function which removed epochs that contained values beyond the ±50 mV limit. The removal of the epochs was to remove both muscle and eye movement artifacts. One algorithmically intensive method to remove ocular artifacts was described in a research experiment dedicated to this task alone (Selvan and Srinivasan 1999) and involved adaptive noise cancellation and adaptive signal processing using recurrent ANN.

A more complete comparison of various noise decontamination methods used on EEG data is shown in Table 16 in Appendix A.

Noise decontamination methods fall into two broad categories – multi-sensor and single-sensor. When multiple EEG sensor applications are needed, independent component analysis (ICA) is a popular noise decontamination method. This research uses a single electrode, and so that category of noise decontamination methods is of interest. Some of the most common single-sensor noise decontamination methods include manual (visual) noise removal, averaging over time, band pass filtering, and more sophisticated methods described above. Since the methods vary in literature this research compares different noise removal methods.

2.4.3 Window Size, Overlap, Envelope, and Normalization

Since EEG is a time varying signal and pattern recognition algorithms seek to classify changes in that signal over time, it makes sense to label stretches of time (sometimes called epochs or windows) with one or more classifications. If an algorithm were classifying train cars as being an engine, passenger car, caboose, etc. – it would be straightforward to select start and endpoints. When it comes to thought processes, however, there is a great deal that is not known regarding what is the optimal window size for different applications. Indeed, a number of other aspects regarding windowing vary from application to application as described in the literature, and there is also little describing why the selection of one set of parameters over another was chosen, or if it was optimal. Some of these parameters include the size of the window (in samples or seconds), the overlapping of windows (which may also imply ambiguity of single classification), and another aspect called the windowing envelope.

The windowing envelope is important because the simple act of taking an EEG signal from one start point to one end point implies multiplication of the signal with a rectangular pulse (being of magnitude 1 within the window, and magnitude 0 elsewhere). Multiplication in the time domain implies convolution in the frequency domain, which may yield undesired ripples of frequencies, unrelated to EEG data being investigated. Other window envelope functions exist, such as a raised cosine (Hanning window). Equation 4 shows the formula for the Hanning window, where $w(n)$ is the Hanning window of size w_{size} seconds for a given sampling rate f_s.

$$w(n) = \begin{cases} 0 & n < 1 \\ \dfrac{1 - \cos\left(n * \left(\frac{\pi}{2} * \mathrm{w_{size}} * f_s\right)\right)}{2} & 1 \le n \le w_{size} * f_s \\ 0 & n > w_{size} * f_s \end{cases}$$

Equation 4 Hanning Window

The Hanning window produces much fewer and lower magnitude ripples, as can be seen in Figure 9.

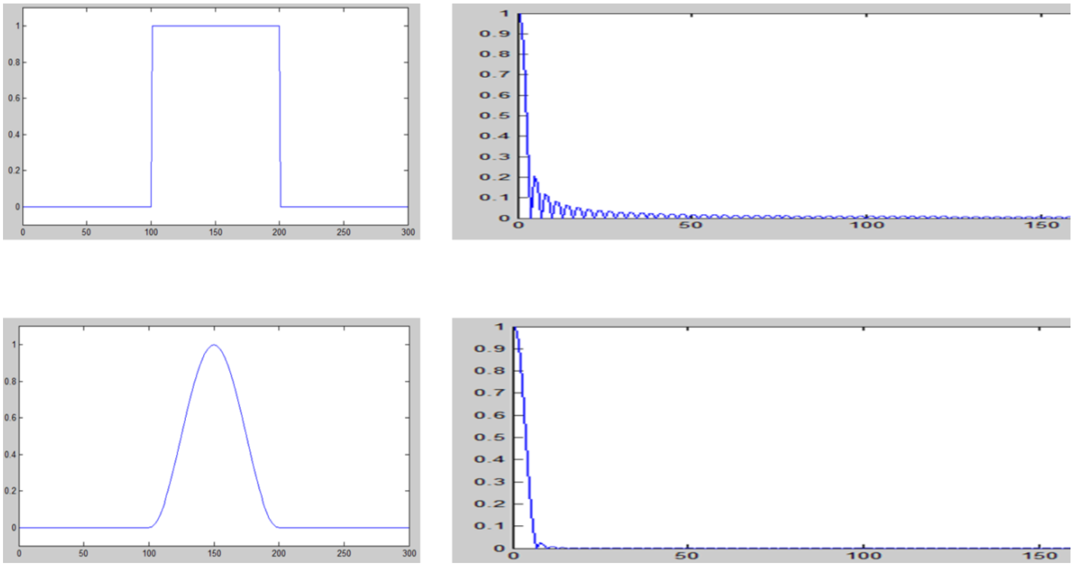


Figure 9 Windowing Envelopes and their Frequency Spectrums

The use of different window parameters can vary greatly. Table 17 in Appendix A shows some examples of the wide variety of windowing parameters described in the literature.

It is also useful to mention normalization at this point, although it reappears in the discussion of machine learning, and in the area of dimensionality reduction. Normalization seeks to remove apparently unimportant amplitude scale information that may otherwise complicate machine learning and pattern recognition algorithms. For example, if the skin or other sensor interface aspect varies from one participant to another, thereby changing the range of voltages recorded by the EEG hardware, an algorithm may look to ignore that overall change, and instead focus on other EEG aspects. For example in (Palaniappan 2006) each individual window was magnitude normalized to values between -1 and 1, and in (Khosrowabadi, et al. 2010) the authors describe normalizing the EEG data for each participant separately. Normalization is also commonly used as a part of machine learning, either normalizing the min and max values of training samples, or normalizing their mean and standard deviations, and then recalling the parameters used for this normalization for use on future, as yet unseen, testing examples. This technique can also be used to normalize individual windows as in the case of (Jung, et al. 1997) where the sample data is first converted to a logarithmic scale and then normalized at each frequency separately by subtracting the session mean and dividing the result by half the difference between the 25th and 75th percentiles of the log power distribution during the session, without memory of the normalization parameters from the training for use the testing samples.

2.4.5 Transformation

Once the EEG data is stored on a computer (both in discrete time – sampling rate, and discrete amplitude – number of bits or resolution), digital signal analysis can convert the signal from a time domain into a more useful domain for analysis and feature extraction. This is done without loss of information as long as certain criteria are met. For example, the Discrete Fourier Transform (DFT), or FFT equivalently, algorithm, can convert a time domain signal into a frequency domain signal. The FFT result yields the same complex valued results as the DFT, except its calculation is optimized for the computer processor. As long as the limitations of Nyquist sampling rate and sufficient resolution of digitization is met, then the digital sampling and subsequent FFT can be accomplished without loss of information (i.e., can be calculated in reverse to recreate the original signal exactly). Other transformations are discussed below.

Before EEG can be used to determine if a participant is paying attention to a video, an extensive amount of signal processing is required, and the literature describes a wide variety of methods to accomplish this. As discussed above EEG signals must be collected via sensors, preprocessing and filtering must remove noise and distortion, and features important to the application are extracted. Expert analysis of EEG data often groups the signals into frequency bands and approximate frequency ranges: Delta Waves are 0 to 4 Hz, Theta Waves (also called Theta Rhythms) are 4 to 7 Hz, Alpha Waves are 8 to 12 Hz, Beta Waves are 12 to 30 Hz, and Gamma Waves are 25 to 100 Hz. It is worth noting that not all the literature is consistent in these band frequencies. This may be attributable to nomenclature differences due to convenience, or to the method of extracting the frequency information particular to that experiment, as is described in more detail in Table 18 in Appendix A, which illustrates for comparative purposes, the differences in the definition of Delta, Theta, Alpha, Beta, and Gamma EEG bands in the literature.

Wavelet and frequency analysis feature extraction methods are popular signal analysis methods (Nagashino, et al. 2002) (Lisetti and Nasoz 2004) (Subasi, et al. 2005) (Palaniappan 2006) (Ito, et al. 2010). The goal of the signal analysis and feature extraction is to create exemplary data (good examples) of factors that define and distinguish the EEG data in terms of what each example classifies. Experiments described in the literature describe methods to create a compressed set of data where it is possible to point to a good example of the EEG features for someone who has one mental (affective) state versus another.

Feature extraction methods for EEG data in literature fall into three major categories:

- Conversion from the amplitude over time into a frequency and phase over time (frequency transform analysis)
- Conversion from the amplitude over time into a scale and position over time (wavelet transform analysis)
- Creation of a spatial-map of how the EEG voltages correlate or do not correlate with each other as spread out over the many sensors placed on the surface of the head.

Since the dry wireless EEG headsets in this research only use one frontal sensor, it is not possible to use them to analyze spatial-map type feature extraction methods such as the "measure of network efficiency" in (De Vico Fallani and al 2008) and finding the 30 most significant pairs of EEG sensors based on their "mutual information" metric (Khosrowabadi, et al. 2010). Nevertheless, these are of interest and may be useful in future research. It is worth noting that the wavelet transform analysis is also concerned with extracting frequency information, described later in the wavelet literature discussion.

Frequency analysis often refers to frequency bands. Banding is a common precedent in the literature. For example (Khare, et al. 2009) discusses collecting FFT information in the Alpha (8-13 Hz) band. (Chakraborty, et al. 2009) discusses using "peak and average" power in seven bands between 0 and 32 Hz. (De Vico Fallani and al 2008) does not take frequency bands but rather it does take the FFT and uses the phase information between different sensors. (Jung, et al. 1997) uses 81 frequency bands from 0.61 to 49.41 Hz. (Nagashino, et al. 2002) uses only time domain information and then presents it to a time delay ANN (presenting ten consecutive 200 samples/sec collected EEG data) essentially training the classifier to act as a frequency analysis method. (Palaniappan 2006) does not use FFT, and instead uses a bank of fifth order band-pass filters to compute the power in each of the delta, theta, alpha, beta, and gamma bands. This implies that the frequency bands are allowed to overlap, due to the dB drop off of the filters. (Ito, et al. 2010) also uses average power spectrum, in this case at 1 Hz intervals from 4 to 22 Hz.

Wavelet transform analysis differs from frequency transform analysis in that frequency transforms convert the amplitude over time data into frequency and phase information whereas Wavelet transforms convert the amplitude over time data into shift and scale information. Nevertheless, the literature suggests using wavelet transform analysis for the purposes of deriving frequency power information. For example, (Subasi, et al. 2005)discussed using Discrete Wavelet Transforms (DWT) to divide the EEG data into four sub-bands (alpha, beta, theta, and delta). Similarly, (S. Lee, B. Abibullaev and W.-S. Kang, et al. 2010) states that DWT was used to compute the "power spectrum features of EEG for each frequency band (alpha, beta, and theta)." A brief description of the DWT can help to explain this seeming contradiction.

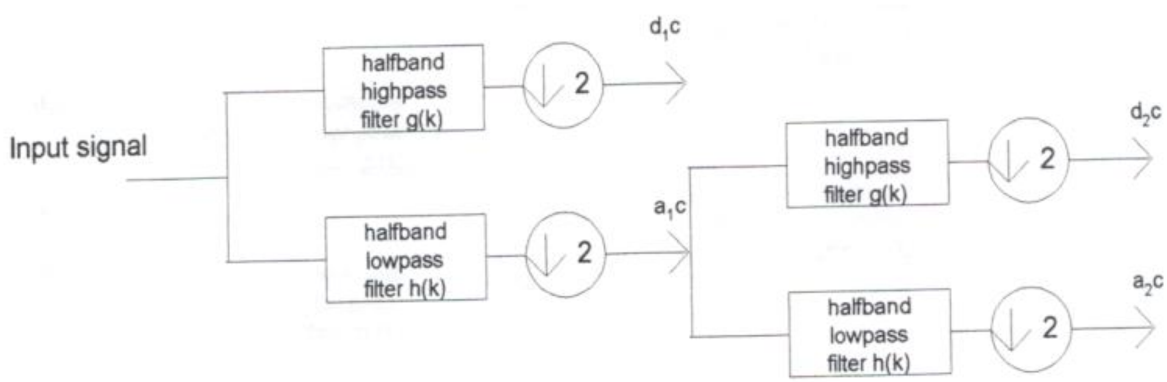


Figure 10 Mallat technique for DWT (Najarian and Splinter 2012)

In Figure 10, the filters h(k) and g(k) are chosen based on the desired wavelet function. The diagram shows a DWT of scale = 2. Notice also that in each step in the scale the algorithm creates a set of approximation coefficients (a_nc) and detail coefficients (d_nc). If the algorithm continues to an nth scale DWT, it will end up with n detail coefficient sets, and n approximation coefficient sets. The size of the sets will be 1/2n, where n is the scale of the coefficients.

The rationale for the use of the DWT for the purposes of frequency band power data computation is perhaps best explained in (Murugappan 2011). Here the paper explains that since the EEG was sampled at 256 Hz, and then the DWT was taken five times, it therefore yielded "D1 (64-128 Hz Noise), D2 32-64 Hz Gamma, D3 16-32 Hz Beta, D4 8-16 Hz Alpha, D5 4-8 Hz Delta, and A5 0-4 Hz Theta." The paper (Murugappan 2011) shows this visually in Figure 11 which details the frequencies for each level in tabular form, showing how different levels of DWT decomposition represent different EEG bands (table is applicable only when the EEG data is sampled at 256 Hz, however a similar band decomposition can be accomplished with other sampling rates that are multiples of powers of two). (Subasi, et al. 2005) explained this using the formula for "pseudo-frequency" which they defined as Fa = (Fc * Fs) / *a*, where Fs is the sampling frequency, *a* is the scale of the DWT, and Fc is the "center frequency or dominant frequency of a wavelet in Hz, defined as the frequency with the highest amplitude in the Fourier transform of the wavelet function." In this article, they use the Daubechies order 2 wavelet function, using simply the fifth scale coefficients.

Frequency Range (Hz)	Decomposition Level	Frequency Bands	Frequency Bandwidth (Hz)
0 - 4	A5	Theta	4
4 – 8	D5	Delta	4
8 – 16	D4	Alpha	8
16 – 32	D3	Beta	16
32 – 64	D2	Gama	32
64 - 128	D1	Noises	64

Figure 11 Mapping of DWT decomposition levels to EEG bands (Murugappan 2011)

Use of these coefficients involves an additional step, that of some statistical measure of those coefficients. For example, in (Murugappan 2011), the D1 coefficients do indeed represent the frequencies from 64 to 128 Hz, but the coefficients themselves are simply a sequence of values. They are extremely phase dependent (i.e. if the EEG data were collected a few milliseconds later, the coefficients would be different) and cannot therefore be used as a feature for input to a machine learning algorithm. Another step must be taken with these coefficients to analyze them and create one or more real numbers (statistical measures) that categorize important features regarding these coefficients. Some of the statistical measures described in the literature include variance, standard deviation, power, entropy (Murugappan 2011), power spectral density (which is equivalent to variance if the mean of the original signal is 0) (S. Lee, B. Abibullaev and W.-S. Kang, et al. 2010), and other non-defined sub-band frequency (Subasi, et al. 2005). For example, if the wavelet decomposition were to be done to the fifth level, then the original signal can be constructed using the wavelet coefficients A5, D5, D4, D3, D2, and D1 – so any of the statistical measures can be used on any of these six coefficient sets.

This concludes the review of related work of signal analysis and feature extraction of EEG biometric data. The literature describes common sensor and digitization settings for alertness measure, but variations exist. For individual sensor EEG data, the use of either frequency band or wavelet parameter data is supported for the creation of a feature vector from the time varying data. Other techniques exist for multiple simultaneous sensor measurements (spatial-temporal analysis). In all of these cases, signal analysis and feature extraction of EEG biometric data is created for the purpose of using the result as an input into a machine learning algorithm.

It is worth noting that one of the primary reasons for converting from the time domain to another domain is to find features that are useful for the classification problem at hand. The brain, for example, can understand the meaning of a spoken word, whether or not the time when the word is spoken has been shifted one way or another by a few seconds. For Machine Learning algorithms to do this, on common processors with limited parallelism, would be computationally prohibitive. For this reason the transformation algorithm is used to convert similar signals, differing only by shift, to be made similar. So essentially, instead of creating a machine learning system which can recognize signals shifted by every increment of time possible, one can simply convolve the EEG signal EEG(t) with a transformation signal Trans(t) as in Equation 5 or cross correlate the signal as in Equation 6 (these are equivalent in the aforementioned examples since the transformation signal is symmetrical/even) with a sine wave or a wavelet and present the resulting magnitude data to the machine learning algorithm.

$$\int_0^t EEG(t)Trans(t-s)ds, \qquad 0 \le t \le \infty$$

Equation 5 Convolution with Transforming Signal

$$\int_0^t EEG(t)Trans(t+s)ds, \qquad 0 \le t \le \infty$$

Equation 6 Cross Correlation with Transforming Signal

In this way, it is possible remove information that is not important to the classification algorithm at hand. The next section will continue the discussion of discarding unimportant data, in the form of dimensionality reduction.

2.4.6 Dimensionality Reduction and Optimal Feature Extraction

As discussed in the previous section, the EEG signal has been transformed into a more convenient form, from the time domain into the frequency/phase domain of the Fourier transform or the scale/shift domain of the wavelet transform. In that discussion it was also noted that some dimensionality reduction took place.

Dimensionality reduction is important for several reasons. First, it greatly reduces computation time for the machine learning algorithms that follow. Second, it helps to avoid the "curse of dimensionality" (discussed further in the next section) which causes over-fitting of training data reducing the generality of the resultant machine learning algorithm. Finally, dimensionality reduction is a means of looking only at the data which is common across many participants and is relevant or common to all those instances in identifying the feature of importance (such as attentiveness), hence for feature extraction. Dimensionality reduction is common in many machine learning implementations, and may include application specific methods, such as banding or Principal Component Analysis (PCA).

PCA is a common dimensionality reduction method which requires samples of data (training data) and once the transformation matrix is calculated, it must accompany the classifier so it can be applied to all new incoming data. PCA does not use any information about the desired workings of the machine learning algorithm, nor the application in question. PCA simply searches for directions in the data that have largest variance and subsequently project the data onto it. (Welling 2012) Another method similar to PCA which seeks to address the classification issue at hand is Fischer Discriminant analysis. Fischer Discriminant analysis seeks to add to PCA, using classification information in the calculation of an optimal linear separation of the data. The Fischer Discriminant seeks to find a linear separation of the data that maximizes the ratio of scatter of inter class data divided by the ratio of scatter of intra class data. (Welling 2012) It is therefore theoretically possible to use the Fischer Discriminant formula (the ratio of inter class scatter over intra class scatter) as the algorithm with which to compare feature extraction methodologies.

The Fischer Discriminant normally seeks to find a linearly transforming matrix, **w**, which maximizes the Equation 7

$$J(\boldsymbol{w}) = \boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w} / \boldsymbol{w}^T \boldsymbol{S}_W \boldsymbol{w}$$

Equation 7 Fischer Discriminant

Since there is no need to look for an optimal **w**, the formula simplifies to Equation 8

$$F = \boldsymbol{S}_B / \boldsymbol{S}_W$$

Equation 8 Simplified Discriminant

Where $\mathbf{S}_B$ and $\mathbf{S}_W$ are defined as the "between classes scatter matrix" and the "within classes scatter matrix" respectively (Welling 2012). Another way of looking at this is using the Davies-Bouldin index which is simply the reciprocal of the proposed Fischer Discriminant Metric F, and also assumes the classifier will use a Euclidean Distance measure of some sort.

When there are only two classes, $\mathbf{S}_B$ can be calculated as the square of the difference of the means and $\mathbf{S}_W$ is calculated as the sum of the within class variances. The scalar value F, which is the ratio of these two scatter matrices, can be used as the Fischer Discriminant Measure for a given feature extraction method.

One example in the literature is (S. Lee, B. Abibullaev and W.-S. Kang, et al. 2010) where a number of wavelet transforms were used and visually compared the fifth level decomposition coefficients of Bior3.1, Sym7, Db4, and Coif5; although an analytic (numerically on its own) comparison was not provided, the paper did compare the feature extraction method in terms of overall system performance. Also important is the work done in (Murugappan 2011) which compared feature extraction methods, largely as part of how they performed as part of the overall system. This paper is important because it discusses an interesting hypothesis of the importance of non-linearity. (Murugappan 2011) notices that the "Entropy" metric seems to be one of the better ones to use and the authors go on to hypothesize that this is because it "…captures the nonlinearity of the EEG signals over different emotions [better] than other statistical features."

Without a complete understanding of the physical brain neuron-firing mechanisms related to when someone is attentive towards a short training video, it would be difficult to make any definitive conclusions as to why linear comparison of feature extraction methods might not reveal the one that will work best in the classifier. Nevertheless, it is well understood that the ANN classifier does allow non-linear pattern recognition. It could be that the ANN is able to solve this pattern recognition problem because it can handle non-linear combinations of input data. Therefore, a purely linear comparison of feature extraction methods might not find the one method that contains the data needed to solve the classification problem the best.

If indeed, linear methods of comparing feature extraction methods are not revealing the best method for feature extraction, then this may imply that common linear methods for dimensionality reduction may flawed. For example PCA is frequently used to reduce a large feature vector down to a smaller dimensionality, such as in (Khare, et al. 2009) and (Chakraborty, et al. 2009). PCA makes assumptions about the ability to find an optimal matrix that is multiplied by the input data. This linear operation, similar to the category using Fischer Discriminant, is by definition taking a linear combination of the total set of data to create a reduced set of data (discarding information). Similarly, the banding methods described in the previous section use linear combinations of adjacent frequency power values to combine them into a single band power value.

If an engineer is looking to reduce the amount of data while minimally impacting classification accuracy, these linear methods may not offer the best solution. It may instead be that linear methods of dimensionality reduction do not yield the best classification accuracy.

In conclusion, dimensionality reduction is important to speed computation, help the machine learning algorithm avoid over-fitting, and extract a subset of the features that are important for the application. Another purpose of dimensionality reduction is to help increase the generalizing abilities of the solution by removing those features that distinguish between participants rather than their mental state. The literature describes several methods of dimensionality reduction including using power (throwing away phase information), grouping frequencies into bands, PCA (Khare, et al. 2009), (Chakraborty, et al. 2009), and converting wavelet coefficients into statistical measures such as entropy and standard deviation (Murugappan 2011).

2.5 Automated Pattern Recognition (Machine Learning Classification Algorithms) of Biometric Data

Even though researchers may have good examples of the data they are looking for, and even though they can extract numerical features from those examples, researchers must also still implement a classification algorithm. Classification algorithms are used to classify the observed signal being a member of one or more categories. (Rumelhart, Widrow and Lehr 1994) Artificial Neural Networks (ANN) have been successfully deployed to perform such tasks as detecting driver sleepiness (Sandberg, et al. 2010) (Hayashi, et al. 2005), and basic statistical analysis tasks such as principal feature classification (Li and Tufts 1997). Other classification algorithms described in the literature include k Nearest Neighbor (kNN) and Support Vector Machines

(SVM). Before going into the detailed review of related work, a brief discussion about training terminology and the ANN follows.

Why is an iterative algorithm that requires training needed in order to perform classification? Why not simply mathematically solve for some conversion formula that takes the input feature vectors and converts them to the desired output (1 for "attentive" and -1 for "not-attentive")? Certainly if a group of exemplary (great examples of) feature vectors **P** is provided where each column of matrix **P** was a single vector, and it is multiplied by some linear transformation matrix **W**, it would be desirable to see that it yields a target vector **T** which is a single column of "correct answers" (1, -1, 1… "attentive," "not-attentive," "attentive"…) for the associated input vector. **W** must be found such that **T** = **WP**. The error of such a system must be as small as possible, so for example, by minimizing the following error function $|\mathbf{E}|^2 = |\mathbf{T}-\mathbf{WP}|^2$. (Hobson-Hargaves 2011) describes that if one could find the inverse of matrix **P** then the problem is simple. Simply let $\mathbf{W} = \mathbf{TP}^{-1}$ and then $|\mathbf{E}|^2$ will equal 0. The problem finding the inverse of matrix **P** requires having a square matrix of great examples of feature vectors (which may not be the case) and also that the mapping from **P** to **T** is linear (which also may not be the case). To solve the linearity assumption, one can use non-linear functions between steps in our algorithm (called "transfer functions" of the ANN layers). To minimize error without the benefit of an inverse matrix, the ANN training methods use iterative training techniques to step by step reduce the error, approaching a solution. The biggest problem with iterative training techniques is that in reducing error the algorithm can "over-fit" or find a solution that only works for the training examples, but works poorly on never before seen examples. This can be exacerbated by having large amounts of data in the form of feature dimensions. The methods described in the literature account for this by using some techniques briefly described here.

(Rumelhart, Widrow and Lehr 1994) ANN programming techniques have evolved over the years, attempting to overcome hurdles in reliable training of the ANN to perform pattern recognition based on example sets. One reliable training method based on multilayer perceptron like devices is known as the back-propagation learning algorithm. The back-propagation learning procedure connects the set of hidden units, to a set of output units, and this input/output pair is called the training set. A target output and an actual output are evaluated using the error correction learning procedure, which in turn leads to measuring the performance and then optimize it, which minimizes the error function.

So looking across the literature, (Rumelhart, Widrow and Lehr 1994) gives an excellent description of ANN types and (Li and Tufts 1997) gives a good example of using an ANN with a winner take all algorithm. Other research (Sandberg, et al. 2010) describes an ANN for EEG data classification that has 3 layers and a single output neuron with 3 or so hidden neurons, and sigmoid activation function (from 0 to 1) to classify driver sleepiness. In yet another experiment (Hayashi, et al. 2005) a 3 layer ANN (6 input nodes, 3 hidden layer nodes, and one output node) is used for each driver. EEG is not used in this experiment (Hayashi, et al. 2005) but instead, Pulse Wave and Steering data were used to detect drowsiness.

There are a number of good examples in literature where an ANN is used specifically in the classification of EEG data. In (Wilson and Bracewell 2002) where the goal is to detect when someone is no longer alert enough to safely operate a vehicle of maintain display vigilance using EEG data, classification is performed using two sets of binary-output multilayer perceptrons. The first set has one output neuron for every spectral band of interest (Alpha, Theta, K, and Delta). Using a three layer ANN, these are trained using artificial data for spectral tuning. The second set are provided with current and past spectral data (from the first set of ANN) to classify into one of seven alertness levels. In (Khare, et al. 2009) the researchers want to use EEG data to classify mental state (e.g., relaxed, imagine moving your right hand, watching a figure being rotated, trivial multiplication, and non-trivial multiplication). The classification algorithm used in (Khare, et al. 2009) two layer feed-forward ANN with 16 inputs, 10 hidden, and one output neuron for each mental task trained to output a 0 for relaxed and 1 for the particular mental task. In (Chakraborty, et al. 2009) EEG data is used to predict the user's facial expression (e.g. anxiety, disgust, fear, happiness, sadness, and relaxation). The classification algorithm used was an ANN learning algorithm with momentum for training the EEG-facial/voice features to the network. After training the ANN is used for mapping EEG space to facial expression space so that given an unknown EEG vector the features of facial expression can be determined by using the ANN. It is worthwhile to note that in this experiment (Chakraborty, et al. 2009) PCA was used on the EEG data to reduce down to a 50 dimensional vector prior to presentation to ANN for correlation to facial data. In (Selvan and Srinivasan 1999) researchers sought to use EEG data including ocular artifacts is used to train an ANN for the purpose of removal of the ocular artifacts (adaptive noise canceller). A recurrent ANN was used to implement the adaptive noise canceller. "To compare its performance, adaptive noise canceller and a cascaded connection of adaptive noise canceller and adaptive signal enhancer are employed. Recurrent ANN using the real-time recurrent learning algorithm (RTRL) is employed for realizing all the [aforementioned] systems." In (Jung, et al. 1997) EEG data is used to predict a participant's alertness and the classifier employed is Back Propagation ANN.

In (Nagashino, et al. 2002) the ANN is trained with 10 input neurons (or more, but without further reduction in error), 30 hidden neurons, and 1 output neuron indicating whether the eyes are open or closed. In (Subasi, et al. 2005) a three layer back propagation network was used with 4 input neurons (one for each band), 10 hidden layer neurons, and three output neurons (one each for alert, drowsy, or sleepy). In (Palaniappan 2006) the researchers used an Elman ANN with single hidden layer trained by the resilient back propagation algorithm.

As mentioned at the beginning of this section, ANN is not the only classification algorithm used in research. One machine learning algorithm is kNN. kNN is supported in the literature and consists of selecting some number k of nearest vectors (using Euclidean distance) from the set of manually classified vectors. The classification of these k vectors is examined, and the most popular classification wins. This method has a number of variations, mostly designed to overcome limitations when the vectors contain many dimensions, or less relevant dimensions whose scale is large compared to the other dimensions, thereby giving excessive non-relevant contribution to the distance measures.

SVM is an algorithm that, like a single layer ANN, can be used to find a line, plane, or hyper-plane that separates the two classes of vectors with as much distance between the vectors and the separator as possible. If the vectors cannot be linearly separated, then SVM attempts to calculate a non-linear transformation (a kernel function) that maps the vectors into a domain where they can be linearly separated (a domain with more dimensions).

Random Forest is a machine learning algorithm based on classification trees. A classification tree is constructed by looking at each individual dimension of the vectors and seeing if it can be used to split the data. This continues on until, a final classification is made (at the leaf of the tree) and several algorithms exist to optimize the creation and pruning (i.e., optimization) of the classification trees. Random Forest takes this one step farther and creates many decision trees (skipping the pruning step), based on randomly selected ensembles of vectors from the original data set (including, sometimes, repeats of the data, sometimes called bootstrap ensembles).

Genetic Algorithms are a machine learning method that mimics natural selection of life based on chromosomes. Each solution has a mother and father, from which it gets a set of chromosomes. The fitness of the offspring is determined (classification error is one measure that is possible) and only the most fit are selected to have the next generation of offspring. In this way a large number of chromosomes (vector dimension values) can be winnowed out to find the best combination to classify the data set.

Some examples form the literature include (Ito, et al. 2010) which used a weighting of individual features (biased Euclidean distance) generated by a competitive "real-coded genetic algorithm" which seeks to genetically find a weighting to maximize the Euclidean distance between vectors from different classes then kNN was used. In (Ito, et al. 2010) the end result was the selection from one of three choices for classification which were "Matches Mood", "Does not match mood", and "Borderline" (or other). In (Khosrowabadi, et al. 2010) researchers documented that they tried kNN and SVM, with the latter performing a bit better. (Murugappan 2011) described the use of kNN with 2 to 8 values for k tried. Seems best results are with 64 electrodes, sym8 DWT (although not much difference), and Entropy as the parameter used and k=6 for the kNN classification algorithm. (Mostow, Chang and Nelson 2011) discussed trained binary logistic regression classifiers to estimate probability of easy or hard based on EEG data. The researchers specifically describe their testing across users (leaving out one user) and across samples for one user (leaving out one sample). Also described were a number of methods used to deal with disparity in number of examples (many more easy examples, many more imaginary word examples) which can plague both ANN and kNN classification methods.

Finally (De Vico Fallani and al 2008) did not use a classifier and simply graphed average efficiency measures for each band, Theta, Alpha, Beta, Gamma - showing more efficiency for forgotten videos. This implies a human classification by visual examination of the graphs. Also (Lisetti and Nasoz 2004) experimented with three different supervised learning algorithms including k nearest neighbor (kNN), Discriminant function analysis (DFA), and Marquardt back propagation algorithm (MBP) for ANN. MBP seemed to do best, and the addition of the feature extraction (min, max, etc…) also slightly improved performance.

So to conclude the review of related work on automated pattern recognition (machine learning) of biometric data, classification algorithms require human supervised training (such as ANN, SVM, and Random Forests) and/or human supervised selection of exemplary feature vectors for each category (such as for kNN). Once this is done, an unsupervised automated algorithm can recognize the pattern, and suitably classify the biometric data.

2.6 Validation

Validation is an important step in the process of creating any engineering system. In the case of signal processing of EEG for the detection of attentiveness towards short training videos, the validation step must measure whether the system created meets the intended goals and needs of the users. The literature review therefore also includes a survey of different validation methods used by researchers. First, a survey of the number of participants used in the study is provided, including the specific application of the research. Second, the validation of the results is examined, including the variety of methods used by researchers.

2.6.1 Number of Participants

A survey of the number of participants in used in various EEG research experiments is shown in Table 19 in Appendix A. Anywhere from under a dozen participants up to 100 participants are used in the research, with the vast majority of research using 10 to 30 participants.

2.6.2 Validation of Results

A survey of the types of validation methods used by researchers is shown below in Table 20. There are four major styles of validation methods used in EEG research, these being, manual, statistical, cross validation, and participant cross validation. Manual validation requires a trained expert to observe the output of the system to determine whether or not it is functioning as intended. The manual method of validation is commonly used when a visual inspection of data is required, or when there may be a wide range of human interpretations of the data (for example, when a trained expert attempts to identify emotion from a participant's facial expression). Statistical validation is when an analysis of variation is performed (ANOVA) where statistical information such as mean and variance are compared against different categories or classes of data. This implies that the final step of machine learning (identification of the category or class from the data) is not performed – there is no automated categorization and only statistical analysis is performed. Another kind of statistical analysis is when two sets of data are compared. This is useful in the case when, for example, and EEG signal has additive noise artificially combined with it, in order to test the performance of a noise removal algorithm. Cross Validation is the validation method that tests the performance of an entire automated categorization system by leaving out random data during training, but leaving some data from every participant in the training set. This is different from participant cross validation where a subset of the participants are removed from the training data set, and then the resulting automated system is tested with those participants who were left out. The latter is perhaps the most rigorous validation method in terms of testing the future performance of an automated system, but nevertheless, the other validation methods are much more popular, and also do reveal important research result information. Table 20 in Appendix A provides a survey of validation methods used in EEG research studies.

CHAPTER 3 Method

3.1 Explanation of the Experiment and Subject Selection Details

The goal of the research was to create a system whereby an observer can employ an easy to deploy EEG device to automatically measure the attentiveness of a participant when watching a short training video. A pilot study was performed to test the equipment and methods (Nussbaum and Hobson-Hargraves 2013). Details of the associated IRB and an explanation of the experiment are in the following sections.

3.1.1 Subject Selection Details (IRB #HM13618)

The EEG data collection uses a product called MindWave Mobile by NeuroSky, a single dry contact EEG device worn over the head like headphones, with a single earlobe ground that connects wirelessly via Bluetooth to recording computer. The device was chosen because it had been proven in similar applications, it was fast for the participant to attach, and did not have large bundles of wires whose movement creates EEG noise artifacts and constrains motion and comfort.

Important for this research is the fact that the identity of the participants is protected. The data from the research is used only for academic purposes and the principal investigator (PI), Co-PIs and graduate students are the only individuals who had access to the data during the study. Records were used whose key identifiers such as names and addresses are removed – after signing the approval form, each participant was given a unique number, and was identified only by that number in all subsequent references.

Collected EEG data, as well as scanned in copies of forms and surveys are saved on a laptop and backup storage with password protections under the control of the investigators.

Permanent storage of the data is on VCU secure servers until it is no longer needed, and then destroyed.

All participants volunteered and read and signed a consent form. The participants were encouraged to ask any further questions and they are also permitted to withdraw from participation at any time, under complete confidentiality. Participants were also told that there will be a short list of questions, then watch a boring video and a short training video during which they will wear an EEG headset to collect the data, and finally another short list of questions will be asked. The participants were reassured that none of the measurements made during the tests will be invasive.

The subject population included both genders with no particular priority given to either, and there was no specific restriction with respect to race, ethnicity or age, except that participants must be 18 years or older. The recruitment of the necessary participants was from within the school itself (students and faculty of ECPI) and the entire pool of day and evening students within ECPI University were given a chance to volunteer for this research. The subject recruitment advertisement was posted on the bulletin boards at ECPI University, and there was no involvement of special cases of subject for this research or others who are likely to be vulnerable were included in this study.

Each subject was selected for this study on a volunteer basis only. The subjects who wished to participate in this study provided their names and signature for the consent form. The consent form was the only research paperwork requiring the participants' actual identification. All of the participants' information, such as the answer to questions and acquired EEG data was labeled only using the identification number.

The participants were free to leave the study at any time and the recruitment of the subject was done in complete anonymity. No one other than those mentioned as a part of this research group have the knowledge of a given subject's participation.

The data acquisition was performed in a setting which included the volunteer subject and the researchers. The subjects participated in this study one at a time.

3.1.2 Explanation of the Experiment

In preparation for the collection of EEG data, a pre-video questionnaire (PreQA), a seven minute video (Video1) and a post-video questionnaire (PostQA) was created. The participant would hand answer PreQA, be fitted with the EEG data collection headset which would briefly be tested, and then the recording of the EEG would commence. The investigator would sit in the room quietly with a stopwatch and notebook to record time stamps of important events as well as any other observations, such as the starting of the Video1 playback, and then the investigator would start the video. Once the video was completed, the participant would remove the EEG headset and answer PostQA.

Video1 has two halves, a boring half followed by an attentiveness inducing short training video half. Specifically, Video1starts with 3.5 minutes of a silent candle burning with title wording stating "Relax and focus on your breathing." Just before the candle video ends, the title wording disappears and then a 3.5 minute long short training video begins. The short training video has the participant follow along and write on a piece of paper, and then fold that paper two times, and then put it in an envelope. The end result is not shown to the participant beforehand, but the task is very simple, and each step is explained, and then modeled (showing the teacher's hands performing the step) one by one. Throughout the training video the teacher makes encouraging statements to help keep the participant on task. Video1 is specifically designed to induce boredom during the first half, and to induce attentiveness towards a short training video during the second half. The nature of the training segment is also designed so that an observer can tell if the participant is on-task.

The PreQA questionnaire was given to the participant prior to watching the video. It asks the participant their age, gender, and whether or not they have been formally diagnosed with a learning disability or with attention deficit disorder by a medical doctor.

The PostQA questionnaire was given to the participant after watching the video. It asks the participant to express in their own words how they felt and what they were thinking at the beginning, middle, and end of the boring segment, and the same questions about the beginning, middle, and end of the short training video segment.

Data collection started on April 14th, 2013. EEG data was collected from 23 participants as they chose to volunteer until May 7th. The committee then discussed the preliminary results, and some changes were made.

After the 23rd participant, the LD and ADHD questions were changed to remove the phrase “by a medical doctor” to more appropriately reflect how individuals are diagnosed/identified LD or ADHD. The new questionnaire (PreQB) was given to all subsequent participants prior to watching the video. In all cases, the entire experiment was conducted regardless of the self-identification of the participant, so that gossip would not spread amongst the student body that one answer or another might change the experiment.

Also, it was decided that two additional videos (Video2a and Video2b) should be created to make the experiment more robust as discussed below. A small group of participants was randomly selected to view one of these other two videos. In all cases, each participant viewed only one video.

Video2a was created to address the notion that perhaps the system would only identify attentiveness if the participant was listening to audio or someone speaking, since the boring candle half of Video1 was silent, and the training half of Video1 had a teacher speaking instructions. Video2a therefore replaced the silent candle boring first half with a boring lecture. The boring lecture had a teacher reading aloud a segment on the philosophy of the history of philosophy, in a dry and monotone delivery which was intended to induce boredom. Just like Video1, Video2a was also a two part video, with the first half 3.5 minutes being the boring lecture, and the second half 3.5 minutes being the attentiveness inducing short training video having the participant write and fold paper.

Video2b was created to address the notion that perhaps the system would only identify attentiveness if the participant was using their muscles and moving (to write and fold paper). To address this, the first half of the video was left unchanged from Video1, with the silent boring candle, but the second half of the video replaced the folding paper training video with a different attentiveness inducing short training video. In Video2b, the short training video segment has the instructor telling the participant to just listen, and every now and then a single digit number will appear and that the participant will have to remember all three of these numbers and report what they were at the end of the video.

The following table summarizes the three videos used in this research:

Video	First Half	Second Half
Video1	Boring Silent Candle	Attentiveness inducing short training video asking participant to write and fold paper
Video2a	Boring / Passive Lecture	Attentiveness inducing short training video asking participant to write and fold paper
Video2b	Boring Silent Candle	Attentiveness inducing short training video asking participant to remember three numbers

Table 2 Three videos used in this research

3.2 Experimental Procedure and Participants

EEG data was collected from 69 participants over the period from April 14, 2013 through June 6, 2013 at ECPI University. Each participant was told how the experiment would be conducted, and IRB matters were discussed (such as safety, confidentiality, ability to opt-out, etc.). Then the IRB approved consent form was signed. All of the 69 participants indicated they wished to continue, and then were asked to answer a short questionnaire (either PreQA or PreQB) relating to their age, gender, whether or not they felt rested and alert, and whether or not they had ever been diagnosed with a Learning Disability (LD) or Attention Deficit Hyperactivity Disorder (ADHD). As discussed in the background section, recent research has demonstrated the effective use of a single dry frontal electrode for the capture of EEG in similar applications. For this research experiment, the MindWave Mobile device from NeuroSky was used. The EEG signal was collected at 512 Hz in 1 second frames which could later be concatenated to any length. An example of the raw EEG data collected, including blinks, or ocular muscle data contamination, was shown earlier in the background chapter in the section regarding EEG noise decontamination.

The participant was then fitted for the dry contact EEG with single left ear ground clip, sitting in front of the data collection and video playback laptop as seen in Figure 12, and the headset was adjusted until a strong signal was generated without interruption, and blinks could be seen on the EEG graph displayed in real-time.

Figure 12 Participant from Pilot Test wearing EEG headset

Once the setup was operational, the recording was started, and the video playback began. Three different videos were used (as described above Video1, Video2a, or Video2b) and each participant saw only one video. The videos are approximately 7 minutes long, with each video starting with a boring segment. After approximately 3.5 minutes of the boring segment, an attentiveness inducing short training video segment started, also lasting for approximately 3.5 minutes for a total of seven minutes to play back the pair of video segments. During the entire time, stopwatch measurements were taken, as well notes of significant events by the investigator sitting slightly behind and to the right of the participant in the same room. The significant events included whether or not the participant was on-task during the training segments of Video1 and Video2a, and whether or not the participant was largely sitting still (no significant muscular activity) during the training segment of Video2b.

At the end of the video, the participant had to sit still for a moment while the EEG data was saved, and the headset removed. After that, a short post-video questionnaire (PostQA) was given to the participant where, in their own words, they provided information on how they felt and what they were thinking at the beginning, middle, and end of the boring video, and the same for the short training video.

In total, 48 participants watched Video1, while smaller populations of 11 and 10 participants watched Video2a and Video2b respectively. Some of the participants had very noisy EEG data and were excluded, and some of the participants self-identified by answering “Yes” to having in the past been diagnosed with LD or ADHD and were excluded. After these participants were removed, there were 40, 8, and 8 participants remaining who watched videos Video1, Video2a, and Video2b respectively, for a total of fifty-six participants. Table 3 shows this information.

Video	Description of video	Questionnaires	Total Participants	Excluded participants (due to self-identification or noisy data)	Remaining Participants
Video1	Candle + Folding Paper	PreQA and PostQA	23	2-self ID + 3-Noisy	18
Video1	As above	PreQB and PostQA	25	2-self ID + 1-Noisy	22
Video2a	Boring Lecture + Folding Paper	PreQB and PostQA	10	2-self ID + 0-Noisy	8
Video2b	Candle + Remember Numbers	PreQB and PostQA	11	3-self ID + 0-Noisy	8
TOTAL			69		56 (40 Video1 + 8 Video2a + 8 Video2b)

Table 3 Participants

During the study the LD and ADHD questions were changed to more appropriately reflect how individuals are diagnosed/identified LD or ADHD. The first 23 participants used the first questionnaire (all with video pair 1) and the remaining participants used the updated questionnaire, so it would be possible to have groups of participants to compare LD and ADHD question answers.

In total there were 24 male and 45 female participants. Their average age was 29.78 with a standard deviation of 9.59, and a minimum age of 18 and a maximum age of 58. Of the entire population of 69 participants, 13 percent self-identified as having been diagnosed with LD or ADHD (8.7% using questionnaire PreQA and 15.2% using questionnaire PreQB).

Prior to the above experiment being conducted on the 69 participants, a pilot study of the equipment and methods was run several months earlier (Nussbaum and Hobson-Hargraves 2013). Based upon the pilot study results some improvements were made to data collection protocol. These included the use of a fresh AAA battery after every two or three participants for the wireless Bluetooth EEG headset, having a more precise time stamping method for start and end of videos and any other EEG data collection events such distractions, and also minimizing distractions by moving the experiment to a quieter and less trafficked section of the building. Also learned from the pilot study were improvements to the videos themselves. The very boring candle video was selected, and also the short training videos were created such that an observer would be able to observe for certain that the participant was on-task, or was sitting still and remembered the three digits. Finally, the pilot study suggested that all volunteers must be told to turn off cell phones and electronic devices during the 10 minutes or so the entire data collection experiment took place.

3.3 Signal Analysis Methods

Signal analysis of the EEG follows several of the methods described in literature, where these methods are compared using both n-fold cross validation and participant cross validation to find the optimal solution for this specific research application. In addition significant novel improvements are achieved in comparison to existing methods.

The methods are divided into specific steps which, largely sequentially, include:

1. Noise decontamination
2. Window normalization, size, overlap, and envelope
3. Transformation into a frequency or scale domain
4. Additional information and dimensionality reduction

3.3.1. Noise decontamination

Noise decontamination methods that are described in literature for single electrode data fall into four major categories. These include visual inspection, filtering, algorithmic noise removal, and epoch averaging. The methods used for this research include comparison of all of these methods, as well as novel methods that add to the current state of the art. There are also methods for multi-electrode noise decontamination (such as independent component analysis, ICA) not used for comparison in this research.

3.3.1.1 Visual inspection (and novel audio inspection)

Visual inspection is where an EEG analysis expert looks at an amplitude over time graph of EEG data and determines that a portion of the data is contaminated by noise (De Vico Fallani and al 2008) (Gargiulo, et al. 2010) (Goldfine, et al. 2011) (MacLean, Arnell and Cote 2012) or individual noisy EEG channels (Gwin, et al. 2010). This data is then manually removed by using a visual cursor to select start and end points of EEG data that is to be removed. The main issue that arises with this type of noise decontamination method is that it is not possible to guarantee that every EEG expert would remove precisely the same data, and therefore the system could yield different results with different operators. Additionally, there may not be an EEG expert available, trained sufficiently to select specific portions of the EEG data recording when the system is used. For these reasons, the current research only permitted the removal of a participant's an entire EEG recording, and not subsets of recordings that require human selection of start and endpoints. The discarding of the EEG data could also be automated, but in either case, it involves the entire EEG data collected from a participant, and not subsets therein.

This research introduces a novel enhancement to visual inspection of EEG data, and that is audio inspection of EEG data. Most visual inspection software programs allow the stretching and shrinking of EEG data on the screen to either zoom in or gain a wider time perspective, respectively, of the EEG data (Epstein 2012). The same up-sampling and down-sampling is used to compare and align the EEG signal with other signals, such as Functional MRI (fMRI) equipment noise (Groullier, et al. 2007). The current research takes these methods a step further, and uses these frequency shifting methods to produce audio signals within the range of frequencies discernible by the human ear. EEG signals in the bands of interest to cognitive research range from 1 Hz to 30 Hz, whereas human hearing frequency range is from 20 Hz to 20,000 Hz. By speeding up the EEG (by playing it back through the computer speakers at multiples of the recorded speed), it is therefore possible for a user to simultaneously look at the short training video while hearing the EEG signal. As a human controlled noise removal method, the audio feedback may allow less experienced users to hear when something is wrong, where they might have had difficulty doing so visually. More importantly, however, after signal analysis dimensionality reduction takes place and it will be very useful if the user can hear whether or not sufficient data remains to discern between attentive and not-attentive portions of a short training video. This is discussed further in the results section of this paper. Removed participants are also re-introduced into the testing set, to see how the system performs with very noisy EEG data, as well as participants who have self-identified as having LD or ADHD.

3.3.1.2 Filtering (and discussion of DFT/DWT filtering)

The next very popular form of EEG noise removal is frequency filtering of the EEG data. This section describes filtering, but also goes on to provide a discussion of DFT/DWT filtering of EEG which is only lightly treated in the current literature and is worth explaining for the benefit of future research. Methods described in literature include hardware and software filters to remove noise. Hardware solutions are commonplace and include hardware implemented notch filters to remove 50 Hz (Europe) or 60 Hz (USA) line noise and hardware implemented low-pass filters to reduce aliasing issues arising from sub-sampling (Nyquist criteria) (NeuroSky Co. Feb 11, 2012) (BIOPAC Systems 2011). This research uses the aforementioned NeuroSky hardware implemented notch and low pass filters. Software filtering also seeks to remove EEG noise, with a variety of filters being employed. Some of these include elliptical (Palaniappan 2006) and Butterworth (Murugappan 2011) Kalman (Chakraborty, et al. 2009), and other notch (Goldfine, et al. 2011) and band pass filters (Belle, Hobson Hargraves and Najarian 2012) as well as keeping only certain frequencies from a Fourier transform magnitude spectrum (Groullier, et al. 2007). The software filters seek to remove electrical interference (notch filters) as well as low frequency muscular and high frequency muscular and ocular noise. This research compared a variety of Butterworth band pass filters, as well as using no filter at all. The digital Butterworth filter is expressed in z (unit of time delay) notation as shown in Equation 9.

$$\boldsymbol{H(z) = \frac{b(1) + b(2)z^{-1} + \cdots + b(n+1)z^{-n}}{1 + a(2)z^{-1} + \cdots + a(n+1)z^{-n}}}$$

Equation 9 Time Delay Notation Butterworth Filter

The above equation can be implemented as a delay tapped algorithm, where z^{-n} represents a sampled data value from n samples in the past. This essentially means that the original EEG signal is convolved with the row matrix **b** and this value is then divided by the EEG signal convolved with row matrix **a**. So, for example, if a Butterworth band-pass filter is designed from fl = 2 Hz to fh = 40 Hz (cycles per second), and the sampling rate of the original signal is fs = 512 Hz (samples per second), then a band pass filter of order n=10 and of pass band range of fl/fs = 3.9 milli-cycles per sample to fh/fs = 78.1 milli-cycles per sample would have **a** and **b** row matrices with individual elements as shown in Figure 13.

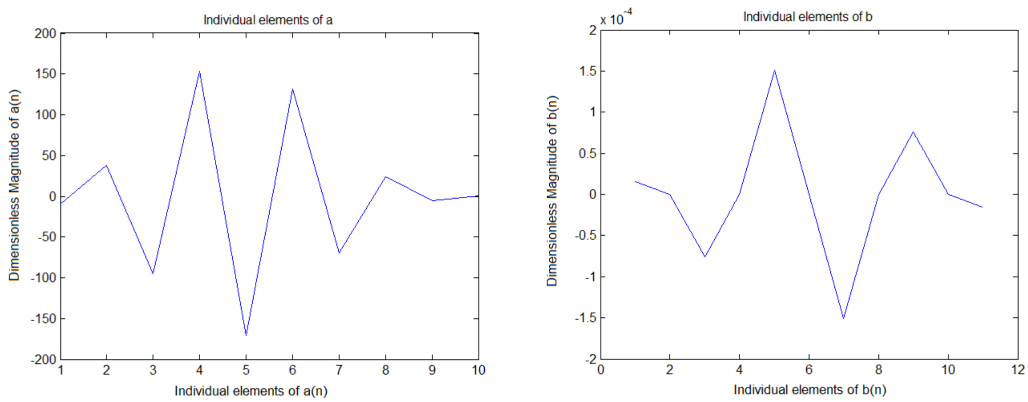


Figure 13 Individual Elements of Denominator a(n) and Numerator b(n)

For the example filter, Figure 14 shows the frequency and phase response.

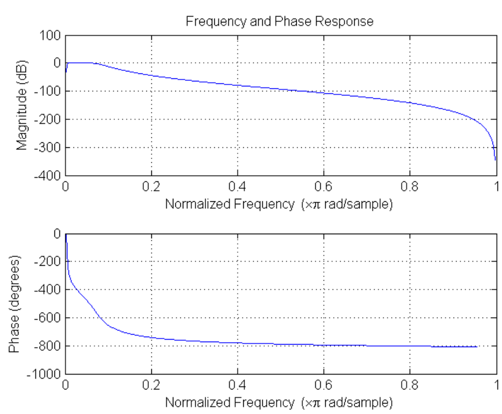


Figure 14 Frequency / Phase Response

The frequency response of this example Butterworth filter has the advantage of a fairly flat pass band of 3.9 to 78.1 milli-cycles per sample, but a slow drop-off of outside the pass band. Nevertheless, the filter is quite effective. Figure 15 shows the frequency spectrum of one second of EEG data sampled at 512 Hz both before and after application of the example Butterworth filter, which proves to be quite effective at band pass filtering from 2 Hz to 40 Hz.

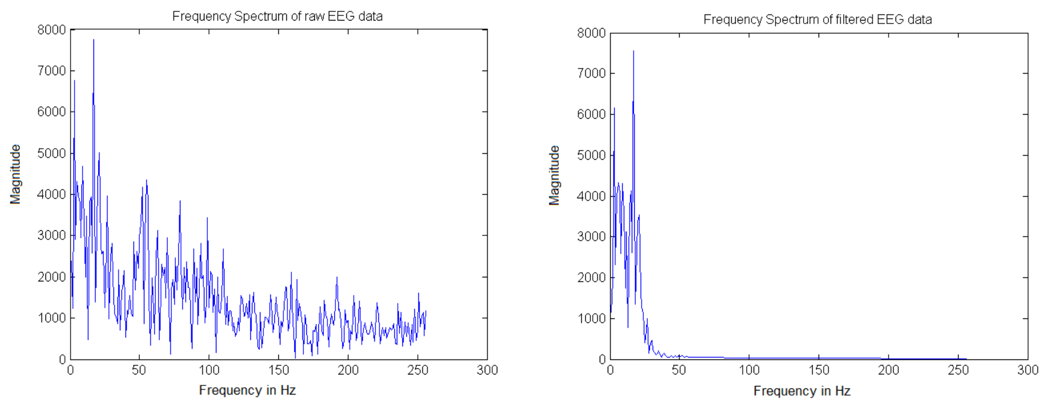


Figure 15 Frequency Spectrum of EEG before and after band pass filter

It is worthwhile to examine the actual EEG signal itself, not just the frequency spectrum, and visually make note of the large amount of potentially useful information the band pass filter may be removing along with the noise. The same one-second of EEG data before and after filtering is shown in Figure 16.

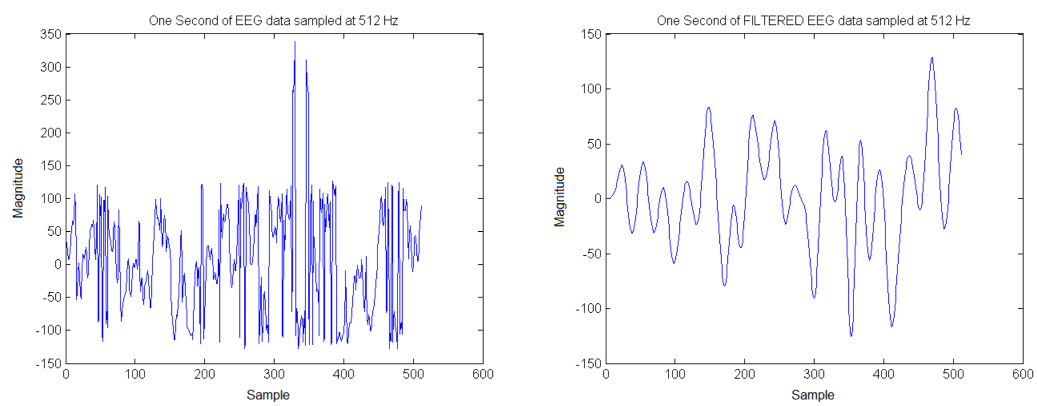
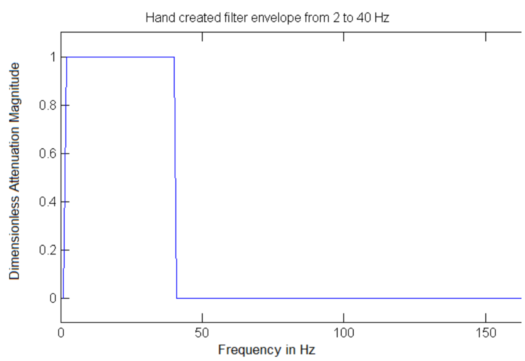


Figure 16 Raw EEG data before and after band pass filter

Note that the use of filtering does not stop at the noise removal phase. Indeed, both the DFT and DWT are essentially frequency filtering methods. The DFT is the method used to arrive at the Frequency Spectrum diagrams shown in Figure 15. It is possible to simply select those frequencies to study directly from the DFT by hand creating a rectangular frequency spectrum envelope from 2 Hz to 40 Hz as seen in Figure 17, and the EEG frequency spectrum after it has been multiplied by that envelope, removing all but 2-40Hz.



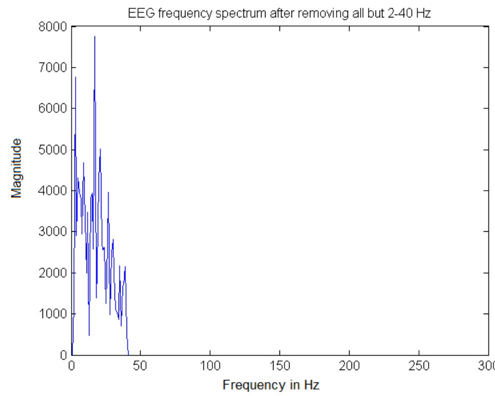


Figure 17 Filter envelope from 2 Hz to 40 Hz, and resultant frequency spectrum

Figure 18 compares the two filtering methods. The figure shows the EEG re-created by taking the inverse transform of the DFT after the high frequency components (>40Hz) have been removed, versus the more traditional Butterworth filter from the previous example. The inverse transform yields similar results, but allows some higher frequency information all the way to 40Hz to be presented without the attenuation of the Butterworth filter drop-off.

This research compared both methods, as well as combinations of the two.

Just as the DFT can be used to filter noise, so too can the DWT. The DWT algorithm convolves a time limited mother wavelet signal, sometimes shaped fairly similarly to the **b** row matrix of the Butterworth equation numerator. Indeed, the DWT Mallat decomposition discussed in the background chapter is implemented with a low pass and a high pass filter that form (along with down-sampling) a single step of the DWT decomposition.

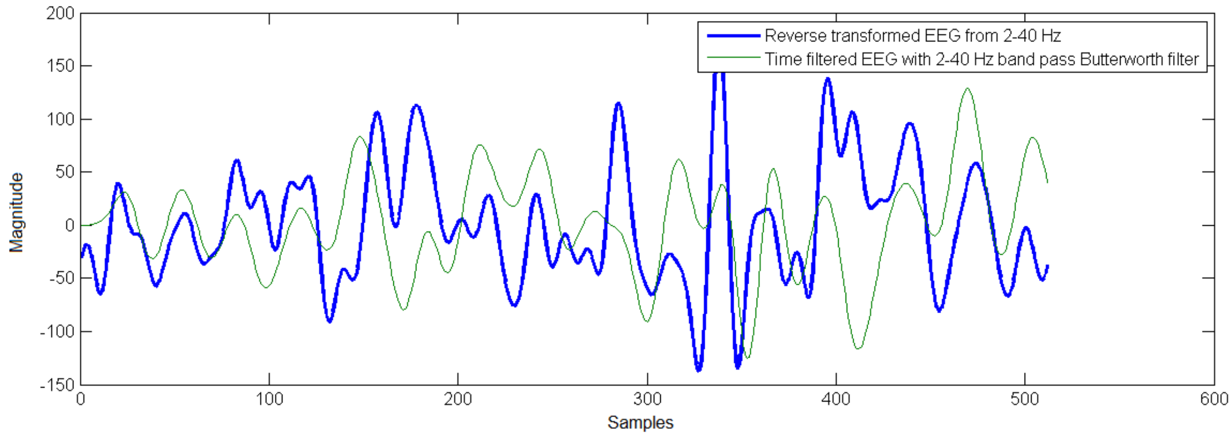


Figure 18 Comparing filtering methods

Since each step of the DWT includes a low pass filter, and then down-sampling by two, the DWT is used to perform frequency analysis on the EEG data many times, and a window size that is a power of two is used to facilitate repeated frequency range halving that is the hallmark of DWT, as was described in the background section and in (Murugappan 2011). Consider the biorthogonal bior3.9 wavelet whose low pass filtering (or analysis filter) length is 20, and high-pass filter (or synthesis filter) length is 4. Figure 19 shows the same one-second of EEG data signal **s**, now filtered by using a bior3.9 six level DWT decomposition, with the decompositions roughly representing the frequency bands $\mathbf{d_1}$ representing 128-256Hz, $\mathbf{d_2}$ representing 64-128Hz, $\mathbf{d_3}$ representing 32-64Hz, $\mathbf{d_4}$ representing 16-32Hz, $\mathbf{d_5}$ representing 8-16Hz, $\mathbf{d_6}$ representing 4-8Hz, and $\mathbf{a_6}$ representing 0-4Hz.

It is possible to get greater granularity of frequency resolution (instead of this $\log_2$ scale of bands), without losing the benefits of spatial resolution gained by the use of the DWT through the use of the DWT packet decomposition. The DWT packet decomposition works just like the Mallat decomposition shown above, which repeatedly performs the low and high pass filtering on the approximation coefficients $\mathbf{a_n}$, but it also performs this on the detail coefficients $\mathbf{d_n}$ as well. This creates a binary tree of wavelet decompositions, which by nature contains redundant information (there is more than one reverse path through the tree to recreate the original signal) but it has greater granularity with regard to frequency analysis.

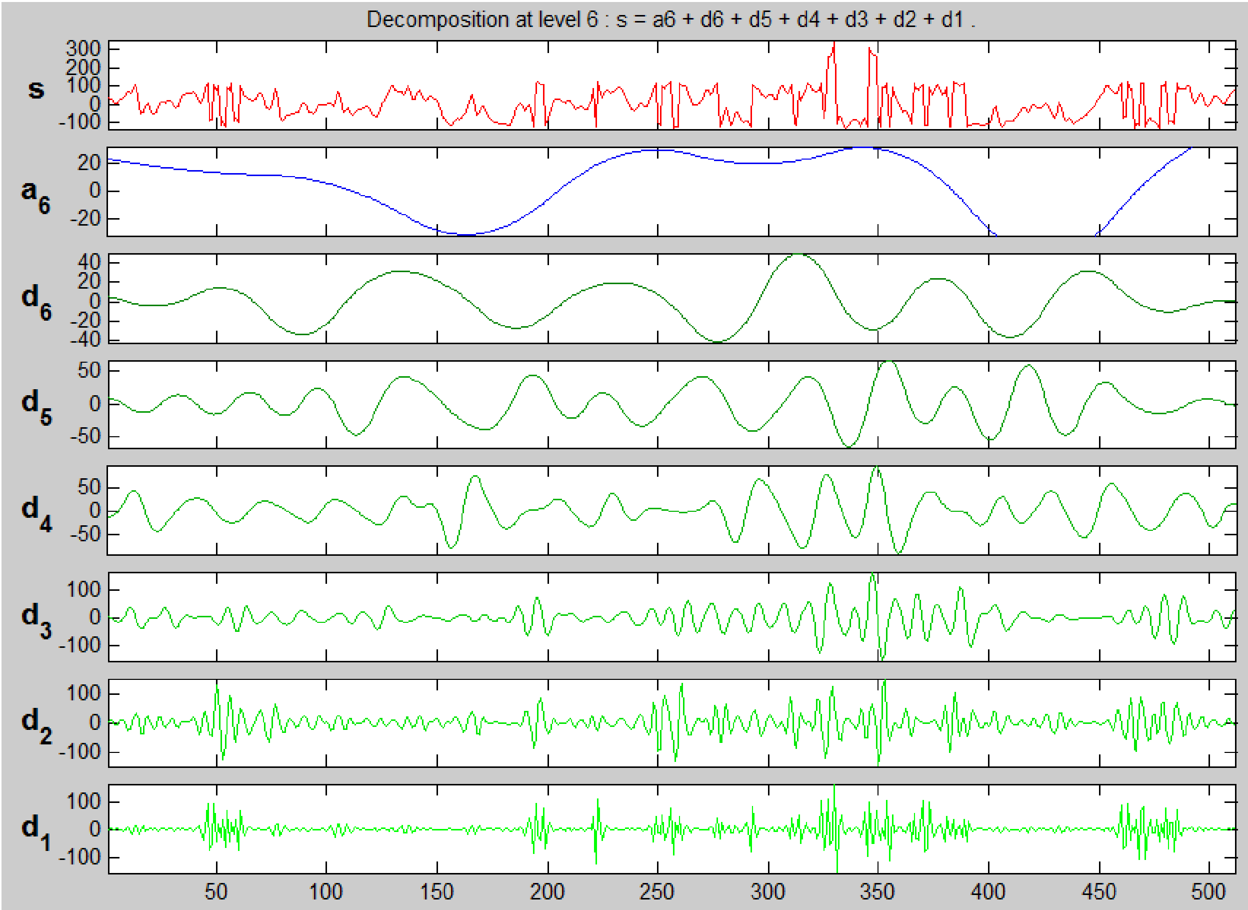


Figure 19 Six level Bior3.9 DWT decomposition representing frequency bands

Figure 20 repeats the bior3.9 sixth level decomposition on that same one-second of EEG data sampled at 512Hz, but this time DWT *packet* decomposition method is used. The results demonstrate the ability to diagram frequency band power both with greater granularity than simple DWT, and also with greater spatial localization than DFT.

The current research compares a number of DWT algorithms, primarily for feature extraction discussed in the transformation methods section below, but it is important to note that DWT transformation implies a filtering and hence a filtering noise removal aspect as well.

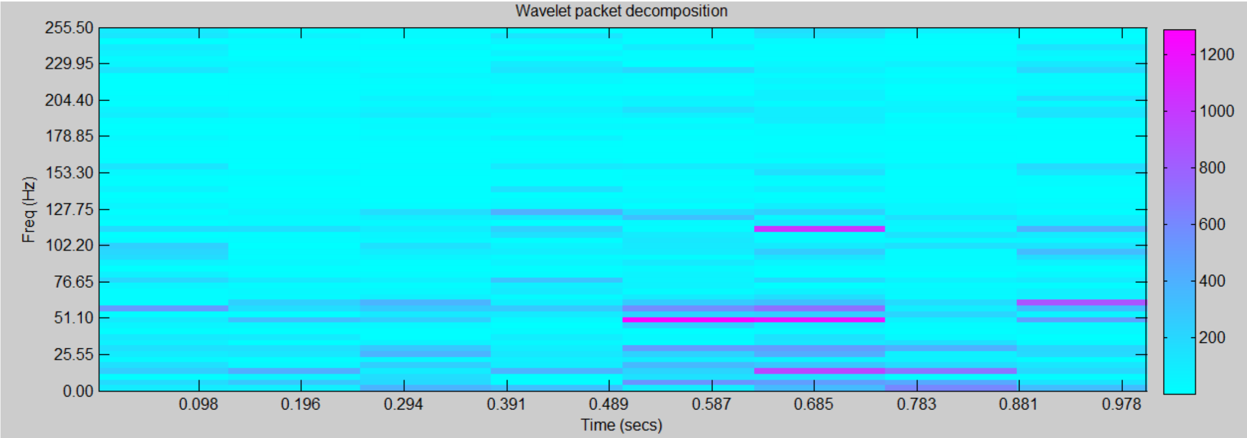


Figure 20 Six level Bior3.9 DWT Packet decomposition representing frequency bands

3.3.1.3 Algorithmic noise removal (and novel threshold selection method)

Algorithmic threshold based noise removal has been the topic of a great deal of recent attention in research particularly because of the large amount of noise added to the EEG signal when the participant is also inside of an MRI machine (Groullier, et al. 2007), (de Munck, et al. 2013) or the use of EEG when running on a treadmill (Gwin, et al. 2010). Even traditional EEG usage with relatively stationary participants not undergoing MRI has received attention in research specifically towards noise decontamination (Ma, et al. 2012) (Safieddine, et al. 2012) (Selvan and Srinivasan 1999). The current research used a wireless EEG headset, and so the noise associated with the movement of the EEG data collection cables (which can be sensitive to the swaying motion created by the heartbeat pulse of the participant) is less of a concern, although the headset can definitely shift and move on the head. Also, the current research used a single electrode, and as such, noise removal algorithms involving Independent Component Analysis (ICA) across electrodes are not applicable. Nevertheless, a number of algorithmic approaches to noise removal based on reducing muscular and ocular artifacts from a single electrode have been researched and were discussed in great detail in the background section. This research compared two of these algorithmic approaches for noise removal, and these were also compared for usage in conjunction with other noise removal methods, and working alone. The first algorithmic method of noise reduction tested as part of this research was the removal of epochs that contain threshold exceeding amplitude levels (where different threshold levels can be selected). This has the disadvantage, in later stages of the process, of making the number of attentive and inattentive examples used for training and testing disparate from one another, and so may also involve techniques such as discarding non-noisy epochs in order to correct this imbalance. For this reason the first algorithm did not produce good results and was no longer used in the research. The second algorithmic method of noise reduction tested as part of this research was the removal of segments of EEG data that exceed a threshold and replacing them with generated data (finding the zero crossings before and after the threshold exceeding data, and replacing that entire range of EEG data with zeroes). This was used in this research, and is referred to as threshold based noise reduction.

As the above descriptions suggest, the selection of a threshold is important and used in several algorithmic noise removal methods described in recent research. The method of selecting the threshold appears to vary from one research description to another. In order to provide additional formalization and justification of the selection method based on the goals it is try to achieve, and this research proposes a novel method not described in existing research. The novel method bases threshold selection on the expected times between noise events, since by definition, the algorithm is trying to remove “infrequent” noise events. The proposed algorithm solves a number of issues. A brief description of three of these issues is described, and then the algorithm for threshold selection is provided, showing that it has the potential to address these issues.

The first issue in threshold selection is that the algorithmic methods of noise decontamination which use a threshold (all of those described above) are doing so based on the assumption that infrequent noise events can be removed, yet the threshold selection they describe does not formally define the expected average time between noise events, nor allow the investigator to adjust the threshold selection based on that definition. Without this ability one cannot know whether or not the threshold selection is very sensitive to the selection of expected time between noise occurrences. Currently described methods of threshold selection use visual inspection, manual selection, and some formal statistical methods (to remove outliers which are more than three-sigma away from the mean, for example) which may assume a symmetrical normal distribution of occurrence of voltage levels.

The premise of symmetrical normal distribution was tested by creating a histogram of EEG voltages. The range of EEG measurements ranging from $\boldsymbol{E_{min}}$ to $\boldsymbol{E_{max}}$ is taken and subdivided into $\boldsymbol{B}$ equally sized bins. For each bin $\boldsymbol{b}$, ranging from 0 to $\boldsymbol{B}$, within a sample set $\boldsymbol{S}$, it is determined if the EEG values $\boldsymbol{E_n}$ fall within the bin's measurement range, $\boldsymbol{C_b}$. Equation 10 shows how the bin counters $\boldsymbol{C_b}$ are calculated.

$$\boldsymbol{C_b = \sum_{n=1}^{S} \begin{cases} 1, & \frac{E_{max} - E_{min}}{B} b \leq E_n < \frac{E_{max} - E_{min}}{B}(b+1) \\ 0, & Otherwise \end{cases}}$$

Equation 10 Multiple Bin Histogram Analysis of Signal

By graphing $\boldsymbol{C_b}$, it can be visually demonstrated that the premise of symmetrical normal distribution is false, as can be seen in Figure 21 which shows an example of the occurrence of various voltage levels over a 7.5 minute EEG recording. Notice the long tail of infrequently occurring voltage levels to the low end of the distribution curve which would sway the median and the mean to the low end. Even the mode is visually susceptible to slight variations in the selected sample, because of the spikey (not smooth) distribution of voltages in the weighted center of the graph. Existing statistical methods do not therefore relate the threshold selected to the tangible desired result of removing noise based on number of occurrences.

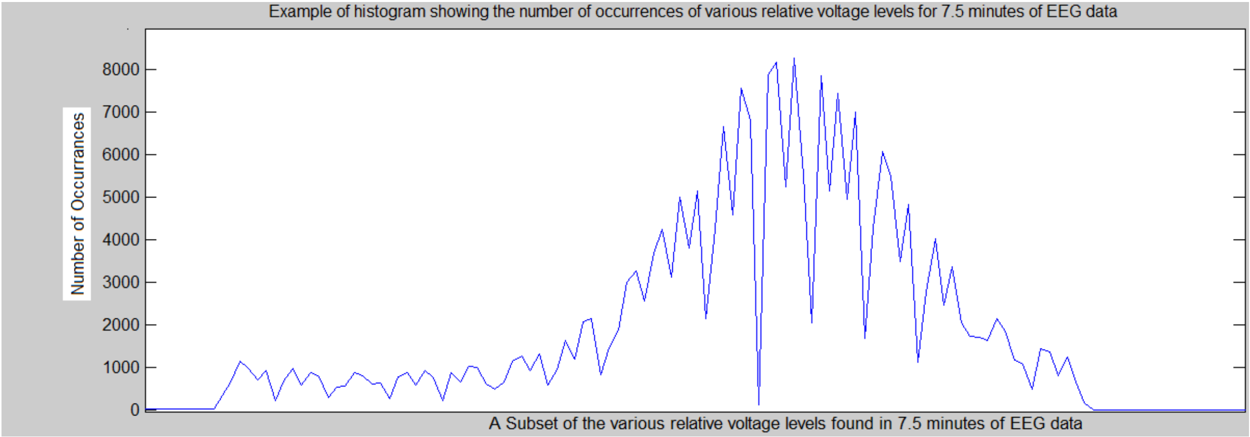


Figure 21 Occurrence of voltage levels over 7.5 minutes of EEG

The second problem is that there is no reason to believe that the threshold useful for positive voltages should be the same as that for negative voltages, and so a problem arises where both a positive and negative voltage thresholds may need to be selected independently. Finally, the third issue is that of calibration. It is assumed that the EEG equipment does not need frequent calibration, and that there is not a problem of slow drift of voltage levels as the equipment is used and ages. Short of developing a model human head which generates calibration voltages, the calibration assumption is cumbersome to verify. Using this novel threshold selection method, the calibration problem is minimized, because the threshold levels selected by the method can be confirmed that they do not drift over time.

The threshold selection method developed in this research requires that there be a number of EEG equipment usage events (in this case, 69 uses over a period of many days, each use generating approximately 7.5 minutes of EEG data). The method also requires that the noise events occur relatively infrequently with respect to the EEG signals being measured. For example, if the participant is constantly shaking their head vigorously, or purposely blinking as fast as they can, then there will be too much noise on top of the EEG signal (indeed, some participants were excluded because the electrode contact was so poor that there was more noise than EEG data present). For this method to be successful, the noise must occur relatively infrequently, and the operator is asked select a time period that is related to how often the noise event is likely to occur. For example, one may want to remove ocular blink noise and may presume that these are events taking place less frequently than once per second on average over the 7.5 minutes of watching a video. The method would therefore require the operator to select one second as the expected time period when on average, no more than one noise event would take place. The operator can vary this time period selection, and indeed the method requires the user to try different time periods in order to confirm that method is robust. As the operator changes the time period expected between noise events, $\boldsymbol{T_n}$ (for example, from 0.5 seconds to 1 second, etc.) the method should arrive at very similar thresholds that are consistent over variations of the expected time period selection. The method of selecting the low and high thresholds $\boldsymbol{T_{low}}$ and $\boldsymbol{T_{high}}$ respectively is shown in Equation 11.

$$\boldsymbol{T_{low} = \frac{E_{max} - E_{min}}{B} b_{low}, \qquad T_{high} = \frac{E_{max} - E_{min}}{B} b_{high}}$$

Equation 11 Low and High Threshold Equations

Where $\boldsymbol{b_{low}}$ is the bin $\boldsymbol{b}$ selected from the ordered set of bins $\boldsymbol{B}$ such that it is the smallest value of $\boldsymbol{b}$ (ranging from 0 to $\boldsymbol{B}$ ordered bins) whose count is greater or equal to the number of discrete samples $\boldsymbol{S}$ divided by the sampling rate $\boldsymbol{f_s}$ multiplied by the expected time period $\boldsymbol{T_n}$ between noise events, as shown in Equation 12. Similarly, $\boldsymbol{b_{high}}$ is calculated as per Equation 13

$$\boldsymbol{b_{low}\ is\ the\ smallest\ b\ where\ C_b \geq \frac{S}{f_s * T_n}}$$

Equation 12 Calculation of Low Bin for Threshold Setting

$$\boldsymbol{b_{high}\ is\ the\ largest\ b\ where\ C_b \geq \frac{S}{f_s * T_n}}$$

Equation 13 Calculation of High Bin for Threshold Setting

Finally, the method also requires that the operator examine whether or not the calculated thresholds are consistent across many uses of the EEG equipment, showing that there is no calibration drift. The method works as follows:

1. The voltage levels are discretely quantified into a number of voltage range sorting bins, with sufficient granularity, and then the voltage levels for each use of the EEG equipment are sorted into these bins. In this research 1000 bins were used for sorting voltage levels, and each 7.5 minute use of the EEG equipment (over several days and 69 participants) was considered a different separate usage of the equipment.
2. Starting from the lowest possible voltage, search increasingly higher voltages until the first voltage is found to have occurred more frequently than the anticipated noise generating events such as blinking, periodic motion, etc. In this experiment, blinking more than once a second averaged over 7.5 minutes was thought to be something that ought not to occur, and so once per second was chosen.
3. The same is repeated to find the upper threshold, starting at the highest voltage and working down.
4. The pair of suggested thresholds is then stored, and the entire process is repeated, once for every usage of the EEG equipment. In this research, 69 uses over 69 different participants were used.
5. Finally, the suggested thresholds are compared. It is expected that there will be infrequent “outlier” suggestions, such as a participant whose EEG collection experienced an unusually high amount of noise, but otherwise, the threshold

suggestions should look fairly consistent and should not show drift over time of EEG equipment usage. Also, variations of the occurrence selection should be tested, and should yield more or fewer infrequent outliers, but not a change to the overall consistency of suggested threshold values.

Figure 22 shows these suggested threshold levels based on a once per second (1/sec) occurrence of noise events. As can be seen from the suggestions, they are rather consistent except for the infrequent outlier, such as in the case of participant 10 which showed greater than 1/sec frequency of occurrence of a very low voltage. Four out of 69 participants is indeed an infrequent outlier situation (5.8%). Additionally, there were a few participants with less noise events than normal, but overall, the threshold level suggestions are extremely consistent with this method, and calibration drift over the several weeks of experiments was not observed.

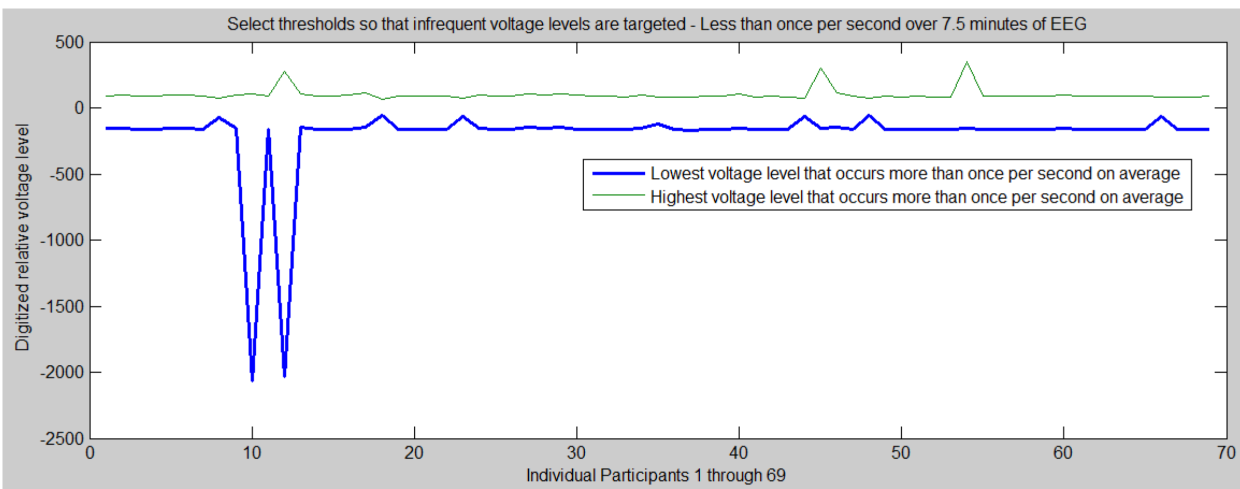


Figure 22 Suggested threshold levels based on expectation of 1 second average time between noise events

Now that equipment calibration and consistency of threshold suggestions has been verified across EEG usages, the method should be tested for sensitivity to occurrence frequency selection. The figures below show that the threshold suggestions remain consistent even when the rate of occurrence values are cut in half or doubled. The only change is the number of outlier situations, but in neither case are there enough outliers to change the strong stable consistency of the threshold suggestions.

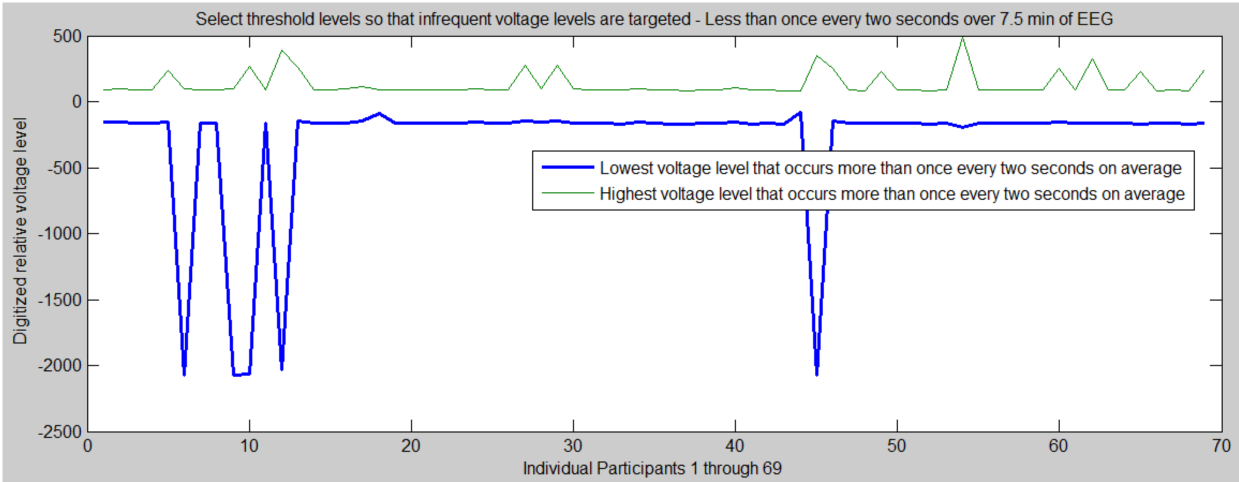


Figure 23 Suggested threshold levels based on expectation of 0.5 seconds average time between noise events

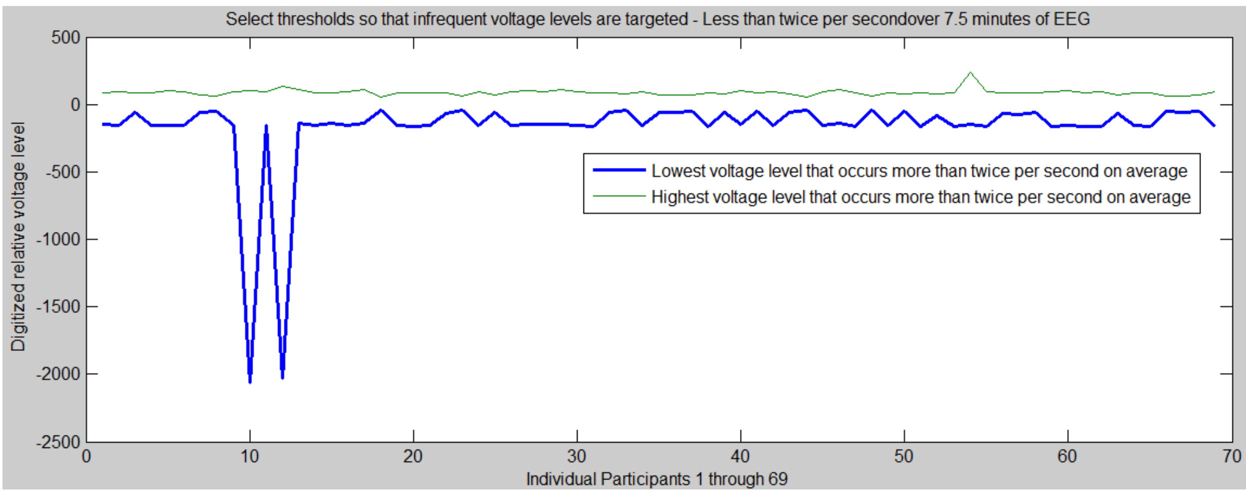


Figure 24 Suggested threshold levels based on expectation of 2 seconds average time between noise events

The above method for threshold selection was used to arrive at separate threshold levels for upper (140) and lower (-170) bounds. These values were then used by the algorithmic noise decontamination methods to remove noise events at the expected time period of occurrence. The current research uses these levels, as well as experimenting with other levels for comparison.

3.3.1.4 Epoch averaging

Even after all of the above methods have been used for noise reduction, current EEG research frequently makes use of epoch averaging as a final measure to provide consistency and to perform noise decontamination. Epoch averaging would only work if the EEG voltage measurements over time were somehow aligned with one another. Instead of averaging time domain data, researchers frequently average shift or phase invariant features that have been extracted from the EEG data. These features are discussed in detail below. Epoch averaging is used in this research after feature extraction has been performed, and a comparison is made against various different numbers of epochs to be averaged. Adjacent (sometimes overlapping) epochs are averaged. Care is taken not to average epochs from different participants, or different situations set up for training of the machine learning algorithm. Also, epoch averaging reduces the total number of epochs available for training or testing. Therefore there is a tradeoff between the ability to smooth out unwanted noise, and keeping important data.

3.3.2. Window size, overlap, and envelope

An extensive discussion of the variations in literature of window size, overlap, and envelope function has already been documented in the background section of this dissertation. The current research compares a variety of these window parameters.

3.3.3. DFT and DWT (and novel methods)

Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT) have already been discussed in detail in the background section of this dissertation, and also above in this section regarding the frequency analysis and noise decontamination aspects of filtering that can take place as a part of these transformations. This research compares both of these methods, and also a number of parameters and options that are available for each of these transformation methods. Two novel methods related to transformation are developed as part of this research.

The first is a method for selecting DWT mother wavelet and decomposition levels. The second is an entirely new transformation method for EEG signals that is based on interdisciplinary learning from control systems. Each is discussed below.

Before describing the novel methods, it is important to note that both the DWT and the novel GF yield a series of coefficients. These coefficients cannot be directly applied to a machine learning algorithm because they are very time shift dependent. In other words, if the EEG signal is shifted forwards or backwards by a few samples, the resulting DWT coefficients will be changed significantly, no matter what wavelet or decomposition level is used. A more phase (or time shift) invariant set of features is needed before presentation to a machine learning algorithm. For this reason, a statistical measure of the coefficients is needed, and several have been proposed. Most recently, a rigorous comparison of such statistical measure has been performed by (Belle, Hobson Hargraves and Najarian 2012) and a selected few have been found optimal for EEG analysis. In this research the use of a different statistical measures were compared and a subset of these was selected for the final system. All of these statistical measures assume an ordered set of coefficients $\boldsymbol{C}$ numbered 1 through N ($\boldsymbol{c_1}$ through $\boldsymbol{c_N}$). The statistical measures compared included Equation 14 Entropy: $\boldsymbol{H(C)}$, Equation 16 Power: $\boldsymbol{P_C}$, Equation 17 Median, Equation 18 Minimum, Equation 19 Maximum, Equation 20 Slope: $\boldsymbol{\mu_{\nabla_C}}$, Equation 21 Max over Mean: $\frac{\mathbf{max_C}}{\boldsymbol{\mu}_\mathbf{C}}$, and Equation 22 Number of Zero Crossings: $\boldsymbol{ZX_C}$.

$$H(C) = -\sum_{i=1}^{N} h(c_i)\, log_2\, h(c_i)$$

Where $\boldsymbol{h(x)}$is the number of occurrences of the value $\boldsymbol{x}$ in $\boldsymbol{C}$, also known as the probability distribution function of $\boldsymbol{C}$ or the histogram of $\boldsymbol{C}$.

Equation 14 Entropy

$$\overline{C_k} = \sum_{i=1}^{N} c_i e^{\frac{-j2\pi ik}{N}}$$

Where j is the imaginary number √-1

Equation 15 Discrete Fourier Transform

$$P_C = \frac{1}{N}\sum_{i=1}^{N} |c_i|^2$$

Equation 16 Power

median$_C$ is the middle value of the set of data ***C*** containing an odd number of values, or the average of the two middle values of ***C*** with an even number of values.

Equation 17 Median

$$min_C = minimum(c_1, c_2, c_3, \dots c_N)$$

Equation 18 Minimum

$$max_C = maximum(c_1, c_2, c_3, \dots c_N)$$

Equation 19 Maximum

$$\mu_{\nabla_C} = \frac{1}{N-1}\sum_{i=1}^{N-1} c_{i+1} - c_i$$

Equation 20 Slope

$$\frac{max_C}{\mu_C} = \frac{max_C}{\frac{1}{N}\sum_{i=1}^{N} c_i}$$

Equation 21 Max over Mean

$$ZX_C = \sum_{i}^{N-1} \begin{cases} 1 & sgn(c_i) <> sgn(c_{i+1}) \\ 0 & sgn(c_i) = sgn(c_{i+1}) \end{cases}$$

Where sgn(x) is the signum or sign function, returning the sign of the value x

Equation 22 Number of Zero Crossings

A variety of these statistical measures were compared for both the DWT and the GF transformation algorithms.

3.3.3.1 Novel method for selecting DWT mother wavelet and decomposition levels

DWT transforms an amplitude over time signal into a shift and scale domain, where a mother wavelet signal $\Psi_{jk}(t)$ is scaled and shifted and then correlated with the signal to be analyzed. The scaling (j) and shifting (k) of the mother wavelet $\Psi_{jk}(t)$ is related to a positive non-zero constant a_0 and the sampling rate of the signal T (measured in samples per second) as in Equation 23:

$$\boldsymbol{\Psi_{jk}(t) = a_0^{\frac{-j}{2}}\Psi(a_0^{-j}t - kT)}$$

Equation 23 Scaling and Shifting of Mother Wavelet

Where $\Psi(t)$ is the continuous time mother wavelet, and the scaling and shifting parameters j and k range from 0 to N-1 and M-1 respectively. The coefficients of the DWT are calculated as shown in Equation 24.

$$\boldsymbol{W_{jk} = \int_{-\infty}^{+\infty} x(t)\Psi_{jk}^{*}(t)dt}$$

Equation 24 Calculation of DWT Coefficients

And the original signal can be reconstructed from the DWT coefficients and the mother wavelet (as well as a constant c based on the selection of the mother wavelet) as shown in Equation 25:

$$\boldsymbol{x(t) = c\sum_{j=0}^{N-1}\sum_{k=0}^{M-1} W_{jk}\,\Psi_{jk}(t)}$$

Equation 25 Restoration of Original Signal from DWT Coefficients

The Quadrature Mirror Filter method of DWT on a discretely sampled signal, such as the EEG signal in this research, requires a pair of filters h(n) and g(n), also derived from the mother wavelet $\Psi(t)$. The DWT algorithm yields detail coefficients and approximation coefficients.

So it is clear that the output of the DWT algorithms is greatly dependent on the selection of the mother wavelet. Also, since the decomposition can be repeated several times/levels (recursively on the resultant approximation coefficients) as described in the background section, the decomposition level greatly affects the output. It is therefore left to the researcher to select the mother wavelet, and the decomposition level, for the application at hand.

This selection of the mother wavelet and the selection of the decomposition level are both unsolved open problems, as noted in the book (Najarian and Splinter 2012). This book also goes on to explain suggestions for making these selections, and confirming the fact that there is no current quantitative method of selection, the research literature provides a wide variety of choices researchers have made, described in the background section of this dissertation. The general rule is that the mother wavelet should resemble the general shape of the signal to be analyzed. Similarly, an intuitive criterion for the selection of the decomposition level is to continue decomposition until the frequencies of interest within the signal are extracted.

This research proposes a method to solve the issue of developing a quantitative method of mother wavelet selection and decomposition level selection for a particular signal and application. The solution is demonstrated on the current EEG signal and application (i.e., attentiveness towards short training videos), and is thought to be generalizable towards other signals and applications.

The solution for DWT mother wavelet and decomposition level selection is described below, both generally, and for the specific signals and application of this research, as follows:

1. Collect multiple data examples of the signal in question from the application in question. In this research, a set of one-second EEG data segments are used, from 69 participants, over 6 different equally spaced time samples, half of which are when the participant is watching a boring video, and the other half when the participant is being attentive towards a short training video.
2. Perform preprocessing on the data examples, just as would be done for the application in question. In the following diagrams, the EEG signal is windowed with a Hanning window, and noise removal algorithm based on thresholds is applied as described earlier. A data example is shown in Figure 25. In addition, the EEG bands are isolated through DFT windowing one by one as described earlier and as shown in Figure 26.

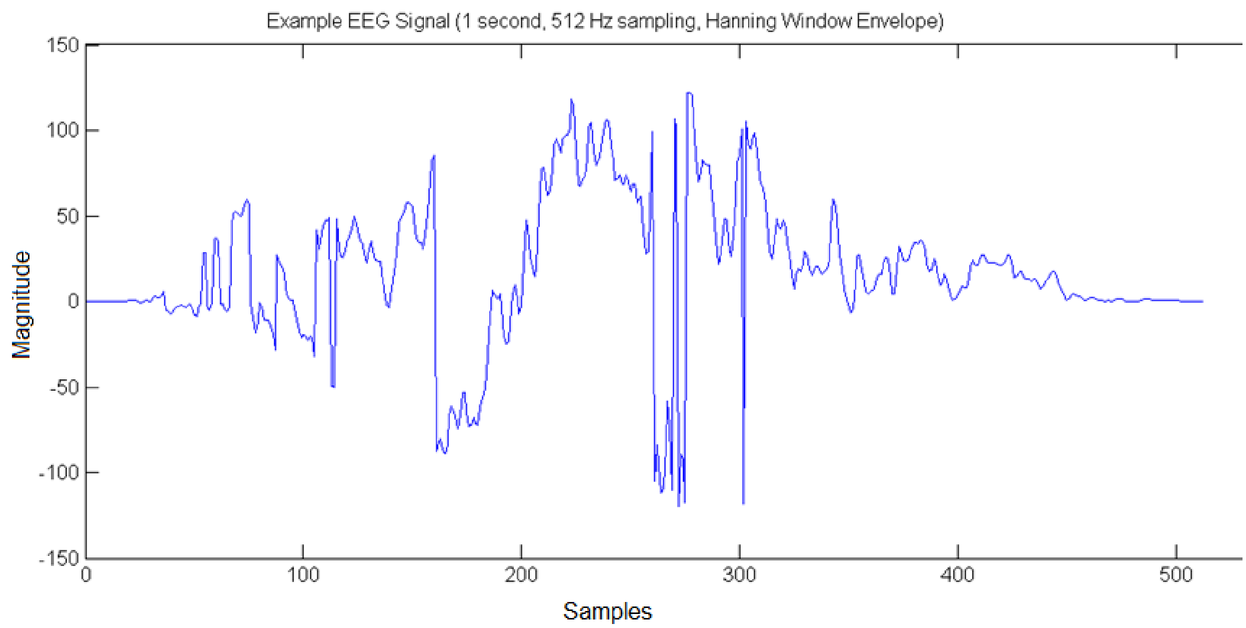


Figure 25 Signal Pre-Processed for the Application

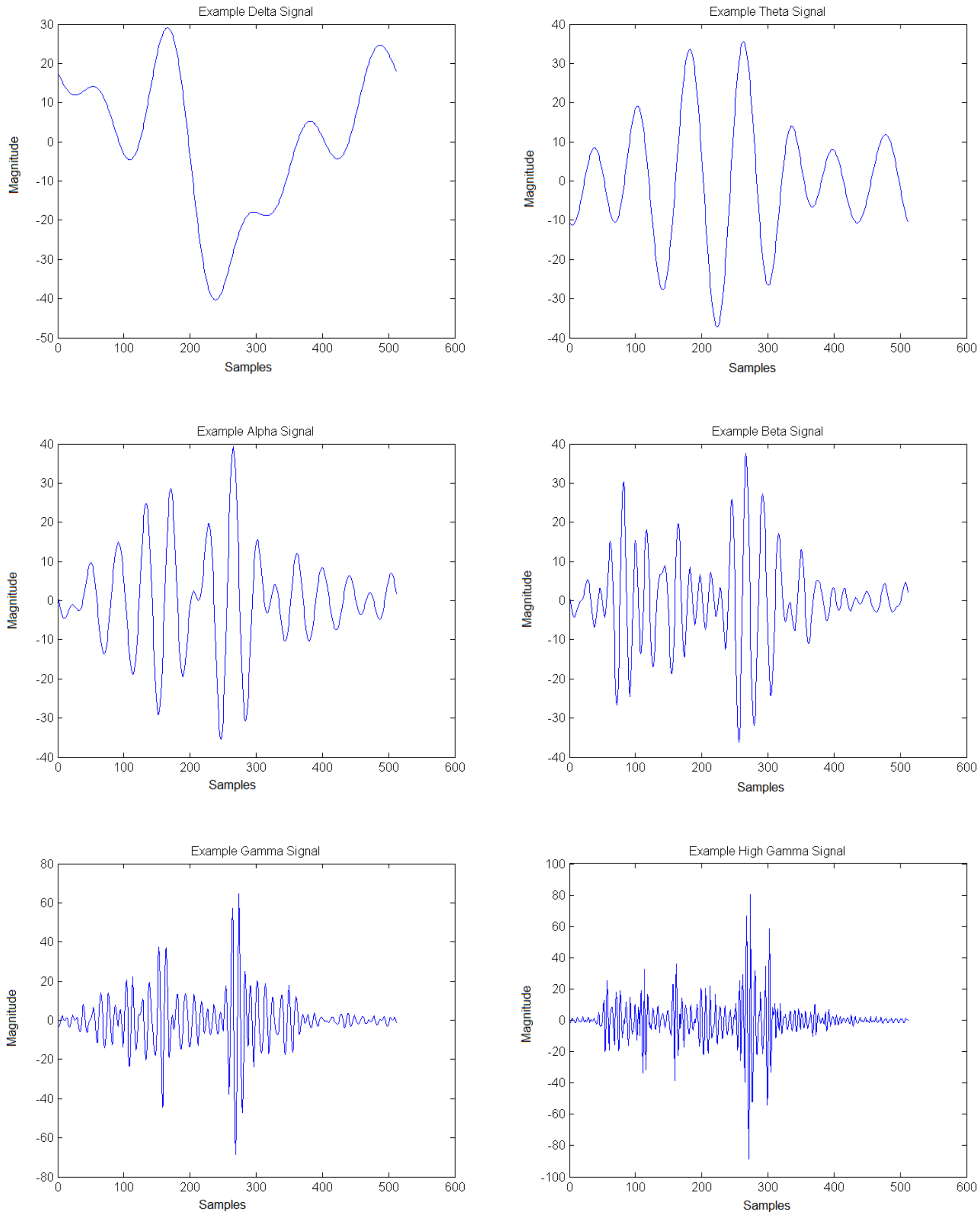


Figure 26 Examples of Extracted EEG Band Signals

3. Across all pre-processed data examples, calculate approximation coefficients for a variety of mother wavelets and decomposition levels. Select a statistical parameter appropriate for the signal and application, and preserve maximal values for that parameter. The current research seeks to select a mother wavelet and decomposition level that most closely approximates the shape of the signal at that

EEG frequency band, at the appropriate scale. When this occurs, the correlation of the mother wavelet at that scale and shift and the EEG signal should be at a maximum, and this is chosen to be the statistical parameter for this research.

4. Average those maximal values across data examples, as appropriate for the application, to arrive at a single maximal value for a given mother wavelet and decomposition level. In this research, the desired goal is to create a system that, independent of the individual participant, can detect attentiveness towards short training videos. For this reason, the maximal values are averaged across the individual participants, and a final single maximal value is used which is that average.
5. Select the mother wavelet and decomposition level that has the highest maximal value according to the selection algorithm described above, but also graph all of the maximal values across all of the decomposition values and all the mother wavelets to check for degenerate situations. These situations may include local maximal values which always occur at the highest or lowest decomposition level, or with many local maximal values. In this research, all but one of the maximal value graphs seem reasonable and as compared to those described in literature as explained in the background section of this dissertation. There does seem to be one maximal value, from the lowest frequency Delta band, that has a local minima that is overshadowed by a trend to get higher and higher values the higher the decomposition level; so the local maxima may be the better choice.

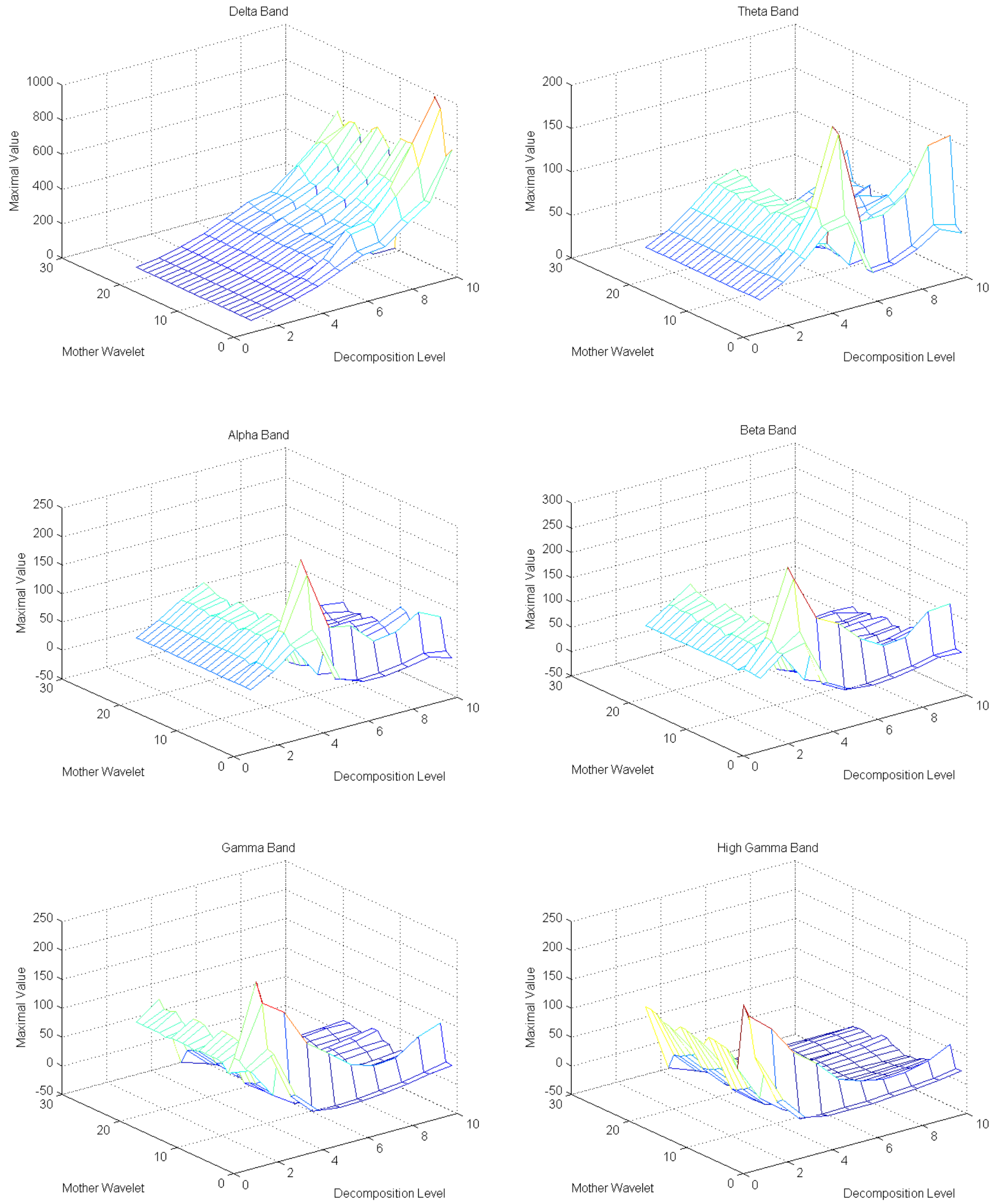


Figure 27 Graphical Results of Novel DWT selection method

The results of the method for DWT mother wavelet and decomposition level is shown below in Figure 27 which illustrates all of the tested decomposition levels (numbered 1 through 10) and all of the tested mother wavelets given by their common abbreviations 'bior1.3', 'bior1.5', 'bior3.3', 'bior3.9', 'bior5.5', 'coif1', 'coif2', 'coif3', 'coif4', 'coif5', 'sym1', 'sym2', 'sym3', 'sym4', 'sym5', 'db1', 'db2', 'db3', 'db4', 'db5', 'dmey' (numbered 1 through 21). Also below are the results shown in tabular format in Table 4.

EEG Band	Frequency Range	Winning Maximal Value	Winning Mother Wavelet	Winning Decomposition Level
Delta	1-4Hz	984	Bior3.9	10 (local max at level 6)
Theta	4-8Hz	247	Bior3.9	5
Alpha	8-16Hz	295	Bior3.9	4
Beta	16-32Hz	273	Bior3.9	3
Gamma	32-64Hz	249	Bior3.9	2
High Gamma	64-128Hz	216	Bior3.9	1

Table 4 Tabular Results of Novel DWT selection method

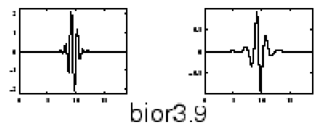


Figure 28 Bior3.9 Bi-Orthogonal compact spline wavelet (image on the left is for decomposition)

3.3.3.2 Novel transformation method.

In addition to the DFT and DWT transformation discussed and compared in this research, the use of a new transformation algorithm, the Gradient Feature (GF) transformation algorithm, was also explained based on cross disciplinary use of the differential or gradient function as a transformation method.

As described above, before DWT detail coefficient or approximation coefficient data can be used as input to a classification algorithm, the series of coefficients must be converted into a handful of descriptive features in the form of statistical measures. In addition, DWT methods frequently used pre-filtering to divide the signal into bands prior to wavelet transformation. GF removes the need for pre-filtering and convolution with a scaled and shifted mother wavelet. Instead GF finds these same descriptive features about the raw amplitude over time EEG data, as well as its derivatives. GF calculates these statistical measures over successive partial derivatives of that same raw data. This lends itself to features which are also computationally simple to extract and making it simpler to deploy on an embedded processing system – which may not have the multiplication speeds of a desktop CPU. Although modern CPU's calculate multiplications as quickly as addition and subtraction, this is not the case for simpler embedded processors – which may benefit from such an algorithm. Future research is suggested towards the optimization of GF to the development of this and other biometric devices.

The operation of the GF algorithm is to take repeated gradients of a series of coefficients ***C*** which represent a time series of EEG samples. The 0th gradient is simply the original signal itself, as in Equation 26

$$\nabla^0 = C$$

Or equivalently:

$$\nabla_i^0 = c_i$$

Where i=1toN, and the individual values $\nabla^0{}_i$ are equal to the original EEG samples c_i.

Equation 26 Zeroth Gradient

Subsequent gradients, like differential equations in the continuous time analogous equations, are recursively calculated,

$$\nabla_i^m = \nabla_{i+1}^{m-1} - \nabla_i^{m-1}$$

Where i=1 to N-m

Equation 27 Gradient of the mth level

Each gradient level yields a series of coefficients ***C*** which are not suitable for presentation to a machine learning algorithm because they are time shift sensitive. As discussed in the DWT section, a number of features must be extracted from the coefficient series in the form of statistical measures. The statistical measures compared included Equation 14 Entropy: ***H(C)***, Equation 16 Power: $\boldsymbol{P_C}$, Equation 17 Median, Equation 18 Minimum, Equation 19 Maximum, Equation 20 Slope: $\boldsymbol{\mu_{\nabla_C}}$, Equation 21 Max over Mean: $\frac{\mathbf{max_C}}{\mathbf{\mu_C}}$, and Equation 22 Number of Zero Crossings: $\boldsymbol{ZX_C}$.

It is useful to note that the GF takes the difference between successive coefficients, which is essentially a type of high pass filter. Lower frequencies with slower moving trends of voltage change have smaller gradients. The slope of the signal is therefore characterized by the first differential. There is an impact of this method on the low frequencies of the EEG signal which influences the overall system performance. Successive applications of the gradient describe the rate of change in the prior gradient, similar to successive applications of a form of high pass filter.

The results section describes the selection of the optimal statistical measures, as well as the levels of gradient used in the final system. It demonstrates that we do not lose important attentiveness indicating information by using the GF method, implying that higher frequencies of EEG are part of the attentiveness measurement equation.

3.4 Machine Learning

A number of machine learning algorithms were tested during the aforementioned Pilot Study, and as such it was observed that some machine learning algorithms performed better than others. Most of the results presented in this research are based on the simplest of these machine learning algorithms, the k Nearest Neighbor (kNN) algorithm, both with and without the use of the novel baseline method, described in more detail below. The results section also compares kNN against other machine learning algorithms, however the methods described below as well as the bulk of the results reported and the description of the novel baseline algorithm are based on kNN.

3.4.1 kNN Algorithm and Variations

There are a number of ways to implement the kNN algorithm but they all fundamentally use a distance measure to find the distances from an EEG data sample to be tested and all the other EEG data samples from the training set. The closest few (the number of these being the value of k) are used to classify the EEG data being tested.

There is no training required to create a kNN machine learning system. The kNN machine learning system consists of a set of $\boldsymbol{Q}$ exemplary vectors $\boldsymbol{M_q}$ each vector having $\boldsymbol{F}$ numerical features as well as one "correct" classification $\boldsymbol{MC_q}$. In this research, the exemplary vectors are the feature vectors derived from the snippets of EEG data extracted from the entirety of EEG data collected (see Figure 29 described in the next section). Preparation of the kNN machine learning algorithm may involve sorting of the exemplary vectors for ease of searching.

As discussed in the background section, there are different distance measures that can be used (Euclidean, Manhattan, and others), and in this research an un-weighted Euclidean distance measure is used. The EEG vector needing to be classified is compared against all exemplary vectors of the kNN machine learning system, as per Equation 28 which describes the calculation of the distance $\boldsymbol{D_{EMq}}$ between an EEG vector $\boldsymbol{E}$ and one of the $\boldsymbol{Q}$ exemplary machine learning vectors $\boldsymbol{M_q}$.

$$D_{EM_q} = \left(\sum_{f=1}^{F} (E(f) - M_q(f))^2 \right)^{\frac{1}{2}}$$

Equation 28 Un-Weighted Euclidean Distance Measure

In some versions of the kNN algorithm, all exemplary vectors are considered together. In this "communal k" implementation, a sorted set exemplary vectors $\boldsymbol{M_q}$ of distances $\boldsymbol{D_{EMq}}$ between the EEG vector $\boldsymbol{E}$ being analyzed and the $\boldsymbol{Q}$ exemplary vectors $\boldsymbol{M_q}$ is calculated is created, as described in Equation 29, and these will be used to determine the k nearest neighbors.

$$D_{EM_q} \leq D_{EM_{q+1}}$$

Equation 29 Sorted Exemplary Vectors for "communal k" kNN algorithm

In another version of the kNN algorithm, which can be called the "individual k" algorithm, the set of $\boldsymbol{M}$ exemplary vectors is divided by their classification $\boldsymbol{MC}$ prior to finding the k minimum distance vectors. So for example, if there are two classifications, $\boldsymbol{M0}$ and $\boldsymbol{M1}$, there would be two sets of exemplary vectors, sorted by their distance measures, each sorted independently, as in

$$D_{EM0_q} \leq D_{EM0_{q+1}} \quad and \quad D_{EM1_q} \leq D_{EM1_{q+1}}$$

Equation 30 Sorted Distance Measures for "individual k" kNN algorithm

In the "individual k" kNN implementation, whichever group of exemplary vectors has the minimum total distance $\boldsymbol{D_{Total}}$ from Equation 31 of the k nearest neighbors, by classification, is selected as the winning classification. So if there are two classifications, 0 and 1, the winning classification would be the one associated with the smaller of the two distances.

$$D_{Total0} = \sum_{i=1}^{k} D_{EM0_i} \quad and \quad D_{Total1} = \sum_{i=1}^{k} D_{EM1_i}$$

Equation 31 "individual k" kNN algorithm calculates total distance for each classification

In the "communal k" kNN implementation, the mode of the classifications $\boldsymbol{MC_k}$ of the k nearest neighbors (the most popular classification) is selected as the winning classification, as described in

$$Winning\ Classification = mode(MC_{1\ to\ k})$$

Equation 32 Winning classification in "communal k" kNN implementation

With the "communal k" algorithm, every classification type may not be represented in the set of k nearest neighbors, whereas in the "individual k" algorithm, there will be at least k nearest neighbors from each classification type.

The value of k can also be changed. As noted in the background section, values of k can vary from 2 to 6 and larger. Also, in most implementations of kNN described in the Background section above, the "communal k" or the most popular classification amongst the k nearest neighbors is the winning classification. To avoid ties, an odd number for k is often employed.

A third algorithm for kNN is proposed for use with the Baseline Algorithm (described in the next section. In this algorithm a variation of the popular "communal k" algorithm is proposed. In the "average of kNN" algorithm, the average of the classifications is used as the winning classification, as per Equation 33. This still requires that the exemplary vectors $\boldsymbol{M_q}$ be sorted by un-weighted Euclidean distance from the EEG vector $\boldsymbol{E}$ in question, and that each exemplary vector $\boldsymbol{M_q}$ is still associated with one single classification $\boldsymbol{MC_q}$, with value 0 or 1.

$$Winning\ Classification = \begin{cases} 1 & mean(MC_{1\ to\ k}) \geq DT \\ 0 & mean(MC_{1\ to\ k}) < DT \end{cases}$$

Equation 33 Winning classification in "average of kNN" implementation

The difference with the "average of kNN" implementation is that the winning classification now depends not only one of two values 0 or 1, but instead on values that can range from 0 to 1. In order to pick a winning classification, there must be a default threshold $\boldsymbol{DT}$ which is used for all EEG vectors being analyzed. The fact that $\boldsymbol{DT}$ can be changed for every participant is the foundation of the Baseline algorithm proposed below.

3.4.2 Novel Baseline Algorithm with kNN

Machine learning algorithms require training, or at least a fixed exemplary set of data points from which distance or linear transformation algorithms are based. Once the entire system has been developed, no further training is possible, because the system will be used by an operator who wishes to determine if a participant is attentive towards a short training video. The operator does not perform manual classification of the participant's EEG data, and therefore needs the automated system to provide this information. Nevertheless, as with other biometric data (such as taking your blood pressure consistently at the same time of day), it is possible for the operator to set a baseline for a particular participant. For example, if the participant is asked to sit still and look at a boring video, then the automated system can use this measurement as a baseline and scale future output of the machine learning algorithm up or down based on that single set baseline. Similarly, a double baseline can be created by having each new participant watch two different videos prior to use of the system. Whether single baseline or double baselines are used, there is no additional training of the system, and no special knowledge by the operator is needed. Nevertheless setting a baseline may be inconvenient, because the system requires this "calibration" exercise to be performed before it can be used to detect attentiveness towards short training videos, however a successfully deployed and usable system of the sort developed in this research will undoubtedly require baseline calibration. This would be similar to a calibration performed by popular voice recognition systems available on the market today, which use calibration to improve accuracy for each new user.

The baseline algorithm, when using the kNN machine learning algorithm is implanted via the following four steps. First, instead of using the mode of the k classifications (the most popular classification amongst the k nearest neighbors) the baseline algorithm uses the mean of the classifications of the k nearest neighbors, or "average of kNN" described above in Equation

33. In this way, the kNN algorithm provides any number of classification levels, depending on the ratio of the different classifications amongst the k nearest neighbors. To determine a classification for an EEG sample under test, a default threshold ***DT*** is used, that is midway between the numerical values used for inattentive and attentiveness classifications. Second, the baseline algorithm requires the use of a larger value for k than traditional values such as 3 or 5. In this research, a value of k=25 was arrived at after analyzing a range of numbers. Third, a baseline is set for each participant, whereby EEG data is collected while the participant is watching what is expected to be a boredom inducing video (if a single baseline is used) or two baselines are set if the participant must watch two different videos, one boring and one which is intended to induce attentiveness (if a double baseline is used). Since it yields the best results and since it is inevitable that the final system will require calibration on a per participant basis, this research presents the results using the double baseline method unless otherwise stated. Fourth and finally, the default threshold is changed based on the baselines for that participant.

So, for example, in Equation 33 without the use of a baseline, one might use ***DT=0.5***, selecting a point midway between the two classification levels 0 and 1. If one does have a baseline of ***R*** EEG vectors from a particular participant that are known come from that participant watching a boring video, $\boldsymbol{E0_r}$, it is possible to calculate a classification $\boldsymbol{C_r}$ for each of ***R*** EEG vectors $\boldsymbol{E0_r}$ and in this way form a baseline threshold ***BT*** as shown in

$$\mathrm{BT} = \mathrm{maximum}(\mathrm{C}_r) \text{ where } C_r = \mathrm{mean}(\mathrm{MC1\ to\ k})\ \mathrm{for\ E0}_r$$

Equation 34 Single Baseline Threshold

One can achieve greater accuracy by using a double baseline. This requires having two baseline sets of ***R*** EEG vectors from a particular participant that is known to come from that participant watching a boring video, $\boldsymbol{E0_r}$, and another set that is known to come from that participant watching an attentiveness inducing short training video, $\boldsymbol{E1_r}$, with the further simplification that the number of EEG vectors $\boldsymbol{R}$ is the same for both. One can calculate a classification $\boldsymbol{C_r}$ for each of $\boldsymbol{R}$ EEG vectors $\boldsymbol{E0_r}$ and $\boldsymbol{E1_r}$ to form a baseline threshold $\boldsymbol{B_2T}$ as shown in Equation 35

$$\mathrm{B_2T} = \mathrm{mean}(\mathrm{C}_r) \text{ where } C_r = \mathrm{mean}(\mathrm{MC1 \text{ to } k}) \text{ for } \mathrm{E0}_r \text{ and } \mathrm{E1}_r$$

Equation 35 Double Baseline Threshold

Therefore the double baseline threshold $\boldsymbol{B_2T}$ is simply the average of the arrived at "average of kNN" algorithm values for each of an equal number $\boldsymbol{R}$ of baseline EEG vectors from the participant watching a boring video $\boldsymbol{E0_r}$ and an attentiveness inducing video $\boldsymbol{E1_r}$.

3.5 Validation of the Proposed Solution

The proposed solution is intended to require no further supervised training. An operator should be able to use the system and method to collect EEG data from a participant who is watching a short training video, and automatically tell if that participant is indicating attentiveness towards that video, through the use of the EEG data alone. The validation of the system is therefore proposed along these same lines.

Several different validation methods have been used in other EEG research. These were described in detail in the background chapter and they include four major styles; manual, statistical, cross validation, and participant cross validation. The last two are particularly useful in testing of an entirely automated classification system intended for real world use, and the first of these two (cross validation) is the most popular in recent research. Nevertheless, participant cross validation is more rigorous and can reveal issues that might take place in real world deployment.

The method used in this research allows for both cross validation and participant cross validation. In addition, the novel method of applying the use of baseline calibration for this application provides improvements beyond the training of the machine learning system.

3.5.1 Overall Measure of Accuracy

A large amount of EEG data is collected from each participant, both while they are watching the boring segment of the video and while they are watching the attentiveness inducing short training video segment. A portion of this EEG data, selected from each of the two segments is used for training and subsequent measurement of accuracy of the system as a whole. Approximately eighty seconds of total EEG data is selected from each participant, half occurring from the boring segment of the video, and the other half from the attentiveness inducing section. For the purposes of machine learning training and measurement of accuracy, these snippets of EEG data are manually classified “correctly” as being inattentive and attentive, respectively.

The figure below graphically shows a storyboard style depiction of Video1. Six screen captures from various sequential points in the video are the shown in this storyboard. In figure 29 below the storyboard, an amplitude versus time graph of the raw EEG data collected from one of the participants is shown. Below the raw EEG graphic the figure shows two smaller snippets extracted from the EEG data, roughly at the location and size of the snippets used for the training of the machine learning algorithms. The majority of snippets used for training are in roughly the same locations for each participant. Some analyses that use all of the EEG data, splitting it down the middle to divide non-attentiveness and attentiveness, are also described in the results section.

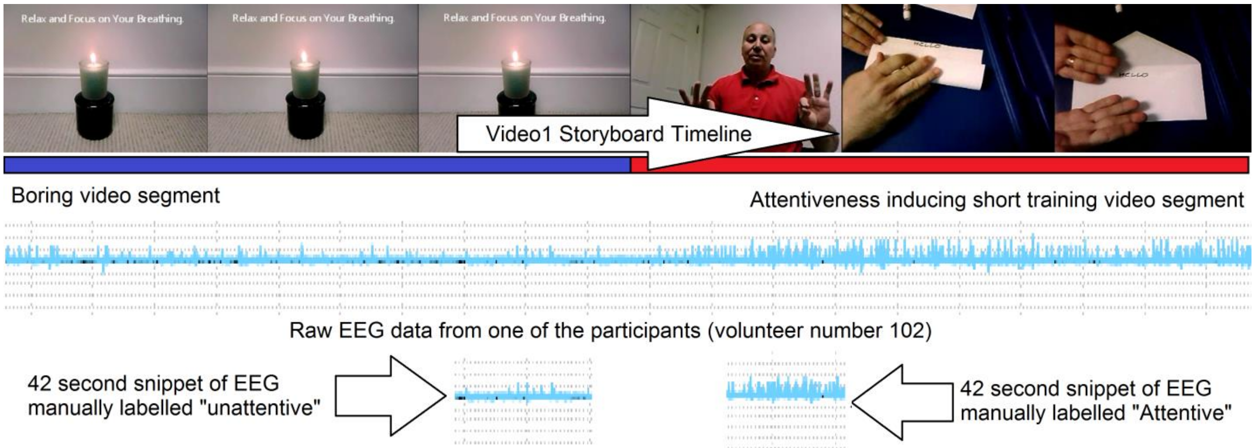


Figure 29 Graphical Representation of EEG snippets used for training machine learning algorithms.

Naturally, it is possible that the manual "correct" classification is actually incorrect such as would be the case when a participant is expressing attentiveness during the boring segment of the video. For example, one participant actually turned around and asked the investigator "Nothing is going to pop up and scare me, will it?" during the boring candle video. One can sometimes identify these situations through the observation notes of the investigator, or from the post video questionnaire where the participants describe in their own words what they were thinking and feeling. This form of triangulation is described in the results section, and it compares the written qualitative data to the quantitative data. This triangulation is not included in the accuracy measurements of the system, because it is not part of the quantifiable EEG data presented to the system, and as such, the manual categories given to the EEG data are assumed to be correct. Therefore it is understood that the system is never intended to be 100% accurate with regard to the indications of attentiveness, because of the complexity of the human mind.

The accuracy of the machine learning algorithm is reported differently; depending on whether or not cross validation or participant cross validation is used. The following sections describe these differences. In all cases, there are an equal number of boring and attentive epochs available for training – so that the system is not incorrectly identifying the percentage of examples in the data set rather than the attentiveness of the participants.

3.5.2 Accuracy Measure for Cross Validation (n-Fold Validation)

When cross validation is used, the pool of EEG data samples from all of the participants is manually classified (as noted above) and then a random subset of these EEG data samples are removed from the pool and not used for training the machine learning algorithm. Instead, this subset is used to test the accuracy of the algorithm. This is repeated n times, and the number of samples withheld from the pool is 1/n percentage of the total number of samples available. This is therefore called n-Fold cross validation, and a popular value for ***n*** in machine learning testing and research is 10, meaning 10 percent of the EEG data samples are left out of training in each of 10 iterations.

Naturally, if 10 percent of the samples are randomly selected to be removed from the training pool, they may come from any of the participants, and either of the classes. Once the machine learning algorithm is trained, the 10 percent of removed EEG samples is used for testing. The machine learning algorithm may correctly classify one of these samples, or incorrectly classify it. The overall accuracy for that iteration of the n-Fold cross validation is equal to the total number true positives plus true negatives divided by the number of samples tested. The overall accuracy is the average of these iteration accuracies over the n iterations.

3.5.3 Accuracy Measure for Participant Cross Validation (LOO Validation)

Participant cross validation differs from cross validation (n-Fold) in that the EEG samples removed from the training set (reserved for testing) come from one or more individual participants. That is to say, the training set does not contain any EEG samples from the removed participants, and the testing set contains only EEG samples from the removed participants. One popular way of performing participant cross validation is by using Leave One Out (LOO) participant cross validation. That means that each iteration removes a single participant's EEG data from the training set, and that is repeated p times, where p is the number of participants.

When using LOO participant cross validation, the machine learning algorithm produces a categorization result for each EEG epoch presented to it. For example, during a 42 second time period, if the window size is 1 second long with 1 second step size, and no multi-epoch averaging, there would be 42 EEG epochs presented to the machine learning algorithm which may produce 42 different classifications. In order to account for the vicissitudes of human thought and the ephemeral nature of attentiveness, the machine learning results are averaged over a defined period of time. In the case of this research, the time periods are those selected from within the boring segment and from within the attentiveness inducing video. Therefore, if there are 42 epochs of EEG data within the selected boring section of the video, the average of the machine learning classification results for each of those epochs is used as the overall classification for that time segment. This research uses a hard threshold for the categorization of the whole time segment, and so in this case the mean of the classifications followed by a hard center value threshold is equivalent to the mode of the categorizations. There are other possible categorization methods, including the use of a "band of uncertainty" where the final classification of the time period is given as inattentive, attentive, or uncertain (if the average classification falls within a central band or range) but this method was not found to improve the overall accuracy of the machine learning largely because when the machine learning got it wrong, it got it very wrong.

So for each iteration of the LOO algorithm, the machine learning algorithm can score 100% (if both the boring and attentiveness sections are correctly classified), 50% if only one is correctly classified, and 0% if both are incorrectly classified. This is then averaged over the p iterations (where p is the number of participants, and therefore the number of iterations when using LOO participant cross validation).

3.6 Final System

The final system is an automated system that uses EEG to provide measurements indicative of attentiveness towards short training videos. A summary flowchart of the steps performed by the final system are shown in Figure 30.

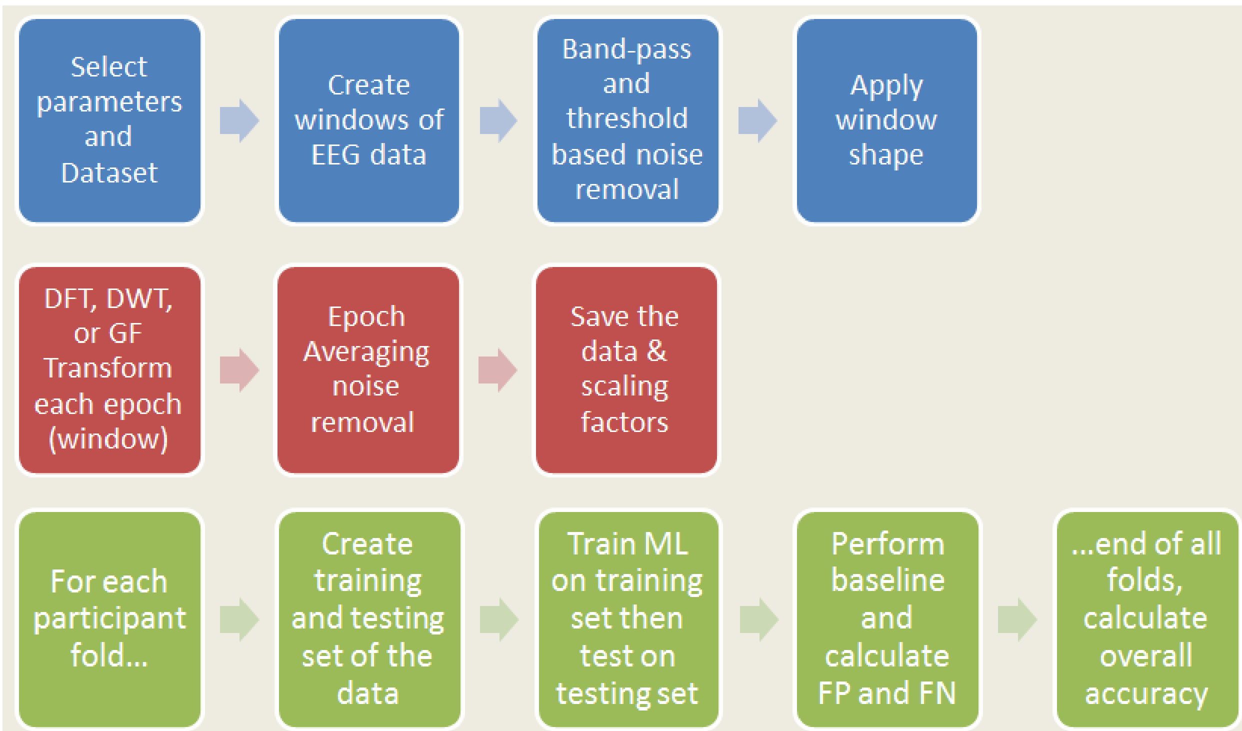


Figure 30 Final System Diagram

The system consists of the aforementioned mechanism for EEG data collection, as well as the signal analysis and machine learning software based on the trained system. For kNN, the system must also include a set of exemplary vectors and their respective categorizations. Finally, if scaling was used during the training, a scaling vector must also be part of the system to scale each new incoming EEG epochs.

Three final systems were created, one each for each of the major signal analysis techniques (DFT, DWT, and GF). At first, all parameters were optimized (through experimentation described in the results section) for use on the EEG snippets taken from the 84 seconds of video from the 40 participants (those that neither self-identified nor had very noisy EEG data). The parameters and accuracy results are shown below in Table 5. The table shows the system settings as they were optimized for participants who viewed Video1.

Alg	Wsz	Wstp	Eavg	BPNR	TNR	Norm	Base	Han	ParSet	Snip	Accur
DFT	1	1	2	no	no	yes	yes	no	40	yes	94%
DWT	1	1	7	no	no	yes	yes	no	40	yes	89%
GF	1	1	2	no	no	yes	yes	no	40	yes	93%

Table 5 System settings optimized for Video1.

The abbreviations in the above table stand for:

- Alg – Transformation algorithm used
- Wsz – window size, in seconds
- Wstp – window step size (distance between starts of windows) in seconds
- Eavg – Epoch averaging noise reduction, in number of epochs averaged
- BPNR – Band Pass (Butterworth 2-40Hz) Noise Reduction
- TNR – Threshold based Noise Reduction
- Norm – Ensemble normalization of each feature to RMS=1
- Base – Baseline per participant in use
- Han – Hanning window shape in used (rectangular if not)
- ParSet – The set of participants (40, 56, or all 69)
- Snip – Whether snippets (84 seconds) are used (otherwise 400 seconds). In both cases, half of the time is boring video (42 or 200 seconds) and the other half is attentiveness inducing short training videos
- Accur – Final accuracy based on % of video segments with less than 50% incorrect classification

These same parameters also yielded good results when EEG snippets from 56 participants were used, including 8 participants who watched video2a and 8 who watched video2b.

Alg	Wsz	Wstp	Eavg	BPNR	TNR	Norm	Base	Han	ParSet	Snip	Accur
DFT	1	1	2	No	No	Yes	Yes	No	DS56	Yes	83%
DWT	1	1	7	No	No	Yes	Yes	No	DS56	Yes	82%
GF	1	1	2	No	No	Yes	Yes	No	DS56	Yes	90%

Table 6 System settings optimized for Video1, then tested on Video1, Video2a, and Video2b

The above results are very good and demonstrate that the solution is robust enough to handle these different cases. Because the end product is intended to be used for short training videos that may or may not require muscle movement, additional tuning of the parameters was performed on the 56 participant data set, and accuracies were improved. These are the final parameters intended for the three systems.

Alg	Wsz	Wstp	Eavg	BPNR	TNR	Norm	Base	Han	ParSet	Snip	Accur
DFT	1	1	2	Yes	Yes	Yes	Yes	Yes	DS56	Yes	85%
DWT	1	1	2	No	No	Yes	Yes	No	DS56	Yes	85%
GF	1	1	2	No	Yes	No	Yes	No	DS56	Yes	91%

Table 7 System settings optimized for Video1, Video2a, and Video2b

Using these new parameters (changes are highlighted in yellow above) provided an improvement in accuracy on the 56 participant snippet data. These are the final settings of the three systems created in this research.

It is useful to note that the system performs well when muscular activity is included in the training video segments (Video1 and Video2a) and also when muscular activity is specifically observed not to take place (Video2b). The qualitative analysis section below and in Appendix C shows that the difference in attentiveness between boring and attentiveness inducing video segments is less when muscular activity is not present, but it is still significant and can be so observed.

CHAPTER 4 Results

A summary of the results is provided, followed by additional detail. Cross validation (10-Fold cross validation, randomly removing individual EEG samples across all participants) results can be as high as 99%. Because of this, the results presented here focus on the more difficult to achieve participant cross validation (LOO). The datasets used are described in Table 8, plus the data set ***DS69*** which was used for both quantitative and qualitative analysis (it not only included every participant, but also incuded all 400 seconds og EEG data rather than the snippets described in Figure 29 Graphical Representation of EEG snippets used for training machine learning algorithms.) The following two tables show the preliminary and final results (Table 9 and Table 10 respectively).

Video	Description of video	Questionnaires	Total Participants	Excluded participants (due to self-identification or noisy data)	Remaining Participants	Short Snippets Included in which Data Sets
Video1	Candle + Folding Paper	PreQA and PostQA	23	2-self ID + 3-Noisy	18	DS40, DS56
Video1	As above	PreQB and PostQA	25	2-self ID + 1-Noisy	22	DS22, DS40, DS56
Video2a	Boring Lecture + Folding Paper	PreQB and PostQA	10	2-self ID + 0-Noisy	8	DS56
Video2b	Candle + Remember Numbers	PreQB and PostQA	11	3-self ID + 0-Noisy	8	DS56
TOTAL			69		56	

Table 8 Participant assigned data sets

Parameters Optimized for Video1	DS40	DS56	DS22
DFT	94%	83%	89%
DWT	89%	82%	93%
GF	93%	90%	89%

Table 9 Table of Preliminary Results

Parameters Optimized for all 3 Video1, Video2a, and Video2b	DS56
DFT	85%
DWT	85%
GF	91%

Table 10 Table of Final Results

Results for DFT, DWT, and GF on the 40 participants viewing Video1 are very similar, using an optimal set of parameters and a simple kNN machine learning with k=25. K is set at 25 so the baseline algorithm can be used by taking the mean of 25 neighbor classes (0 or 1), using the participants threshold to classify the EEG sample, and then choosing the most popular result over the EEG snippet time period which contained approximately 40 seconds of EEG epochs selected for training and testing.

LOO accuracy as high as 89% to 94%, showing ready for to go to market usage.

LOO accuracy does not vary significantly with the addition of both Video2a and Video2b (the 56 volunteer data set) indicating these algorithms cross audio and muscular usage boundaries. LOO accuracy goes down but stays as high as 90%. This supports the evidence that the algorithms are applicable whether or not the boring video has audio and whether or not the attentiveness inducing training video has muscular usage. Additional parameter optimization improves the accuracy on the 56 participant data set as described in the methods section and in further detail below

LOO accuracy goes down somewhat with the removal of the Video1 volunteers who used the old questionnaire (40 volunteers becomes 22). This is likely a reflection on LOO accuracy due to the number of volunteers in the training set. Nevertheless, the accuracy stays as high as 89% to 93%, however this is using different algorithm parameters than with the 40 participant data set. Nevertheless, 89% and higher accuracy proves out the algorithms even when fewer volunteers are available for training, and that the change in the questionnaire did not greatly impact the results of the research.

Novel DWT wavelet and decomposition level selection methods yield good results with 89% accuracy on the 40 participant data set, going down to 85% accuracy on the 56 participant data set with additional tuning of parameter selection. Specific testing on the impact on final accuracies due to the variation of wavelet and decomposition level are presented in detail below and supports the described method of optimal selection.

Traditional DFT also performs well. Accuracy levels of 94% were achieved on the 40 participant data set, with accuracies going down to 85% using the 56 participant data set (Video1, Video2a, and Video 2b).

Refinements to the GF algorithm now make it one of the highest scoring algorithms with an accuracy of 93% on the 40 participant data set, and 90% on the more general 56 participant data set. This is accomplished with the early version of the GF algorithm using up to the 5th derivative. Using the a later faster and more streamlined GF algorithm (needing only the 1st and 2nd derivative) and additional tuning, the GF algorithm provides 91% accuracy on 56 participant data set, with details are provided below.

The described method of noise removal threshold setting is supported by the favorable results of the algorithms above that make use of these selected thresholds across all three videos (56 participant data sets) with results as high as 91% accuracy. In addition, comparison of other thresholds (wider, narrower, and shifted threshold levels) is described in detail below, and shows that when the threshold levels are changed from the optimally selected ones, accuracy goes down.

As noted above, three systems were created, one each for each of the major signal analysis methods. Testing of the three final “Products” (using optimal parameters for DFT, DWT, and GF respectively) was performed on a much wider range of the video (400 seconds versus 84 seconds) and also on all 69 of the participants, regardless of whether they had earlier been excluded due to self-identification or because of a large amount of EEG noise. Here too, the data was equally divided between boring and attentiveness inducing training videos (200 seconds each). Quantitative analysis of accuracy was performed similarly, where the system had to identify EEG epochs within the respective half of the video correctly at least 50% of the time to be marked as successful. One would expect the accuracies to go down dramatically, because now the entirety of the video is being used, including areas where the participants minds are wandering and not necessarily only focusing on the video (see the Triangulation section for more detailed analysis). Nevertheless, these quantitative accuracies were all very good, with DFT and DWT performing at 85% accuracy, and GF providing 81% accuracy.

A closer examination of the participants who influenced lowering the overall accuracy measurement (ex: because they had “false” positives of attentiveness during the boring video segment) were then triangulated with other qualitative data that was collected during the EEG data collection. For example, participant 13’s EEG data was classified as very attentive during the boring part of the video – but the timing data collection notes of the investigator reveal that during the boring video half, the participant turned to the investigator to ask “Nothing’s going to pop out at me, is it?”

Other quantitative information is collected as well during the use of the final system. Ensemble averaging of participants who watched the same video is used to help graph attentiveness over time across many participants at once. Pseudo-audio allows the system user to hear attentiveness change in synch with video, both as a reverse transformation of extracted features, and audio sound representing the ensemble attentiveness measure for that video. These results are described in greater detail below.

4.1 More Detailed Examination of Cross Validation

To perform a more detailed examination of cross validation and participant cross validation, a simple version of the solution was created which did not use baseline optimization. Here a DFT algorithm was used that employed a window size of 5 seconds and window step size of one second. A Butterworth band pass filter from 2 to 40 Hz was used for noise removal, as well as an 8 EEG epoch averaging after the DFT transformation. A raised Cosine Hanning window was used before DFT and only frequency spectra from 2 to 40 Hz was used for input to the machine learning, which was kNN with k=5 (mode of the 5 resulting classifications is used). This method yielded 81% accuracy with LOO participant cross validation across the 40 Video1 volunteers.

In order to more closely examine the impact of cross validation schemes, the above setup was used with a variety of cross validation methods. First, a variety of cross validations where a random subset of the EEG samples, across all participants, is removed from the training set and used only for testing. The most common of these is the 10-fold cross validation, but also compared were 40-fold, 8-fold, and 4-fold. Contrasting with these, a similar set of participant cross validation tests were performed, where a number of EEG samples from one or more participants were removed from the training set, here also using 40, 10, 8, and 4-fold cross validation. The results are summarized in Table 11 below:

N-Fold Participant Cross Validation	% Correct	N-Fold Cross Validation by mixed samples	% Correct
40-fold participant cross validation, Leave One Out (train on 39 volunteers and test on one - rotate through all combinations and average results)	81 %	40-fold Cross Validation (train on 97.5% and test on 2.5% of randomly selected epochs from the set of all 40 volunteers)	99%
10-fold cross validation by VOLUNTEER (train on 36 and test on 4 volunteers - rotate through 10 different combinations and average results)	76%	10-fold cross validation (train on 90% and test on 10% of epochs)	98%
8-fold cross validation by VOLUNTEER (similar to above except train on 35 test on 5)	76 %	8-fold cross validation	98%
4-fold Cross validation by VOLUNTEER (as above)	76%	4-fold cross validation	98%
2-fold cross validation by VOLUNTEER (as above, train on half the participants, and test on the other half)	74%	2-fold cross validation	97%

Table 11 More Detailed Comparison of Cross Validation Methods

The comparison in Table 11 suggests two things. First, the Participant cross validation is more rigorous and difficult in that it reveals lower accuracy results than the comparable cross validation using samples randomly selected from all participants. Second, the selection of LOO participant cross validation is a good selection for overall system analysis, since it does not vary in a very large way from other participant cross validations (6-9% change from LOO accuracy) and it also reflects the real world situation of a finished system being tested on one new participant.

4.2 Epoch Averaging Comparison

As described in the methods section, epoch averaging is a noise reduction method that takes place after the signal has been transformed and features extracted. Each EEG epoch of features is averaged with nearby epochs to smooth out noise. This is particularly effective with the DWT transformation and feature extraction methods in that the noise smoothing can increase overall system accuracy.

Nevertheless, there are tradeoffs with the use of epoch averaging. Equation 36 illustrates this tradeoff. If a set of new averaged EEG epoch features $\boldsymbol{EEG_{fnew}}$ is to be created from an existing set of EEG epoch features $\boldsymbol{EEG_f}$, then for every epoch feature at position ***p***, it is possible to average a range of feature values from positions ***p-Eavg*** through position ***p***, where ***Eavg*** is the number of epochs being averaged. The same holds true for any sequence of EEG epochs so averaged, such as if the epochs averaged occur after the position ***p*** as shown in Equation 37.

$$EEG_{f_{new}}(p) = \frac{\sum_{n=p-Eavg}^{p} EEG_f(n)}{Eavg}$$

Equation 36 Prior Epochs Averaging

$$EEG_{f_{new}}(p) = \frac{\sum_{n=p}^{p+Eavg} EEG_f(n)}{Eavg}$$

Equation 37 Post Epochs Averaging

The tradeoff arises as ***Eavg*** becomes a larger and larger percentage of the total number of epochs available in the snippet. For example, if the snippet under consideration contains 42 epochs, and the averaging window size ***Eavg*** is 7 epochs, then 7 of the 42 epochs can no longer be used (either the first 7, the last 7, or a few at either end of the snippet). This means that less than 84% of the available EEG epochs can be used. There are some methods to avoid this situation by using a weighted average of the EEG epochs at the edges of the snippet. Equation 38 shows one such method of weighted averaging of epochs. In this case, post epoch averaging is being used, and the epoch in question at position p is near to the trailing edge of the snippet, within ***Eavg*** epochs of the last epoch of the snippet. In this the last ***Eavg*** epochs are added along with some weighted amount ***w*** of epoch ***p*** which is then divided by ***Eavg+w*** yielding the weighted average.

$$EEG_{f_{new}}(p) = \frac{w * EEG_f(p) + \sum_{n=q}^{q+Eavg} EEG_f(n)}{Eavg + w}$$

Equation 38 Weighted Epochs Averaging

This was attempted as part of this research but did not improve the overall accuracy of the solution. This is attributable to the fact that the edge epochs using Equation 38 differ from their neighbors because the formula for their calculation is different. If the value of ***w*** is made to be large, the more they resemble un-averaged epochs, whereas if ***w*** is kept small then the edge epochs all become nearly identical, and pad the training set with not useful data.

Figure 31 shows the results of epoch averaging without the use of weighted epoch averaging at the edges of the snippets. All three transformation methods used threshold based noise removal and kNN machine learning with k=25. Also included is a graphed line labeled "vectors/epochs" which graphically demonstrates the epoch average ***Eavg*** size tradeoff.

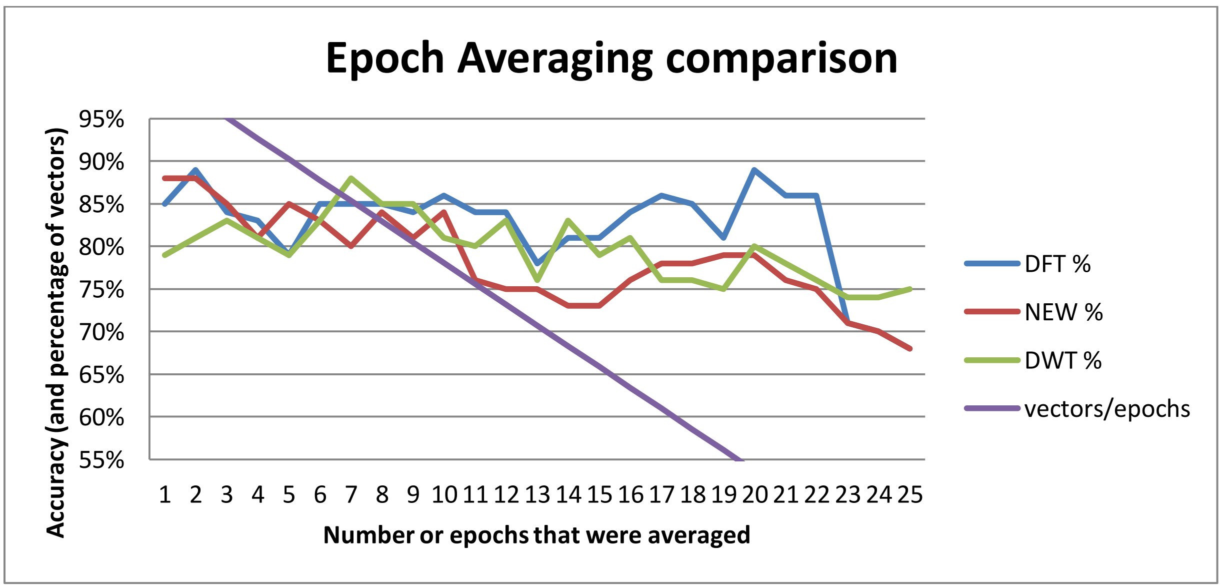


Figure 31 Epoch Averaging comparison

As can be seen with DFT, an accuracy of greater than 85% can be achieved with a variety of ***Eavg*** values, but as values get higher, the total number of averaged vectors available for input into the machine learning algorithm, as a % of originally available EEG epochs, gets lower and lower. For the DFT algorithm, there is no point, therefore, in using an ***Eavg*** value of 20 epochs to get an accuracy of 89% when it is possible to get that same accuracy using an ***Eavg*** value of 2 epochs, and keep a larger percentage of the data available for machine learning.

The resultant optimal values for ***Eavg*** used is therefore 2 epochs for DFT and GF (New) algorithms, and 7 epochs for the DWT algorithm. This yielded accuracies of 89%, 88%, and 88% respectively, using the interim parameters set up for this analysis.

4.3 Window Size Comparison

The size of the window of EEG data, in seconds, is varied and the accuracy resulting values are compared. The experimental setup is the same as described in the above section, using the aforementioned optimal epoch averaging values determined for each transformation type.

The following Figure 32 shows the results of this analysis. Note that the number of training vectors is not reduced by a multiple of the window size, since the window spacing between the start of the windows (the window step size) is kept fixed at one second. By having the larger windows overlap in this way, the reduction in the number of resultant vectors is only similar to the impact of increasing epoch averaging size as described in the previous section.

As can be seen from Figure 32, the DWT algorithm can take advantage of a larger window size of 6 seconds, whereas the other algorithms of DFT and GF (the NEW algorithm) do not benefit from window sizes greater than 1 second (non-overlapping windows). Upon closer examination, the benefits of improving DWT accuracy by 3.4% by using six second window size, over the 88% accuracy achieved by using the 1 second window size were deemed not to outweigh the reduction in the amount of training data available. This decision was reinforced by the seemingly chaotic nature of the DWT accuracy graph in the window size range of 5 to 10 seconds.

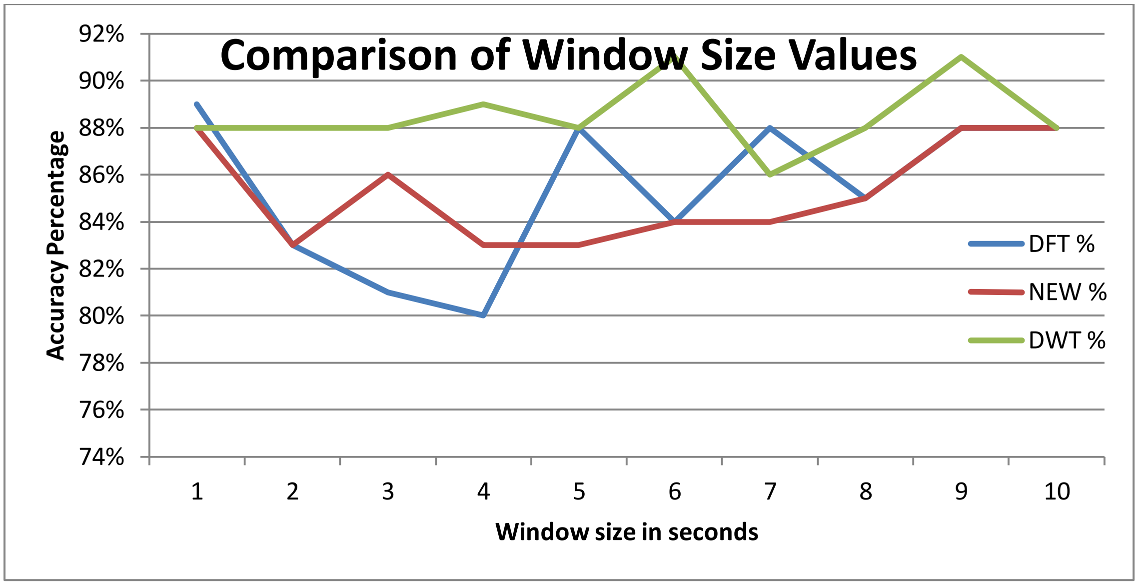


Figure 32 Comparison of window size values

Therefore the optimal window size for all three algorithms (DFT, New/GF, and DWT) are all deemed to be 1 second with 1 second step size (non-overlapping) with resultant accuracies of 89%, 88%, and 88% respectively, using the interim parameters set up for this analysis.

4.4 Switch Setting Comparisons

Some of the settings described in literature which were to be compared in this research are binary in nature. Like a switch, the method is either used or it is not used. Some of these switches control the use of algorithms which may have parameters that can be tuned over a range of values. For example, Threshold Based Noise Removal is an algorithm that can be used, or it can be turned off. If it is used, however, there are high and low threshold values that can be tuned over a wide range of values. This tuning is discussed in a separate section of this dissertation, and for the purposes of this switch setting comparison, the noise removal technique is either used or it is not used.

Five switches are compared in every combination, and the accuracy of the total solution is generated based on the use of one of the three transformation algorithms, the previously described optimal values, and the use of EEG snippets of data from the subset of 40 participants who watched Video1.

The setting of the switches can be thought of as a binary counter counting from decimal value 0 to 31, or 00000 to 11111 in binary, equivalently. The position of the switches in this binary counter analogy are shown in Table 12 and the resulting system accuracy is shown in

Figure 33 Algorithm usage accuracy comparison . It is clear that the baseline method (most significant bit in the counter analogy, and shown in the values 17 to 32 in the figure) has the most significant overall impact on overall system accuracy, while other switch settings have a less significant impact.

Baseline Method used	Normalization of Features	Band Pass Filter Noise Removal	Threshold Noise Removal	Hanning window shape
0	0	0	0	0
0	0	0	0	1
…	…	…	…	…
1	1	1	1	1

Table 12 Algorithm usage switches

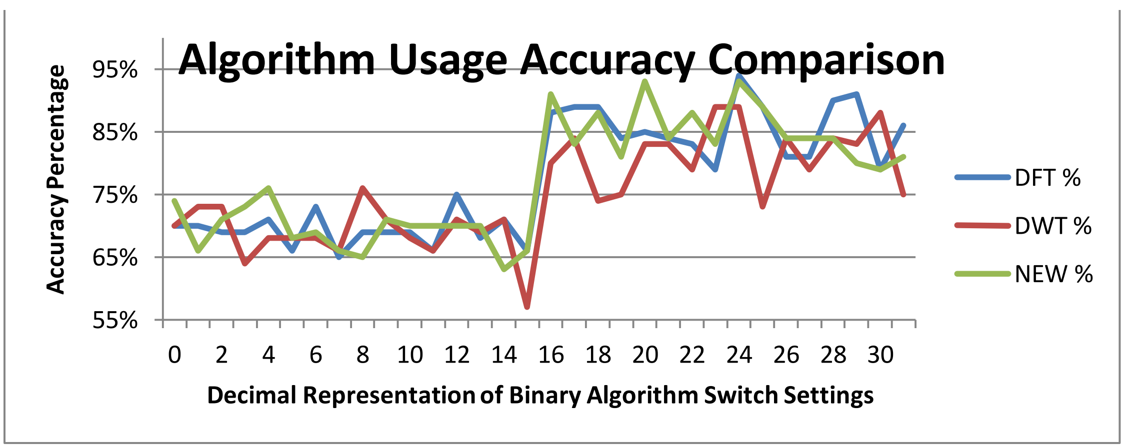


Figure 33 Algorithm usage accuracy comparison

The final optimal selection of switches were described in Table 5 System settings optimized for Video1. These settings yielded the best results for all three transformation algorithms (DFT, DWT, and NEW/GF) of 93%, 89%, and 93% respectively at switch combination 24 (when Baseline and Normalization are on, but other switches are off). Once the number of participants was increased to include participants who watched the other videos, the results were still good, as described earlier in Table 6 System settings optimized for Video1, then tested on Video1, Video2a, and Video2b. The accuracy results were now 83%, 82%, and 90% for the three transformation algorithms, DFT, DWT, and NEW(GF) respectively. This proved out the extensibility of the system for use when the boring segment of the video includes someone giving a boring lecture, and also when the attentiveness inducing short training video was not including muscular movement instructions to the participant (and the participant was observed to be sitting still).

Additional tuning was performed using all three videos to yield the final parameters used in the end systems from this research. These were described in the methods section in Table 7 System settings optimized for Video1, Video2a, and Video2b. Using these final parameters, the three transformation algorithms DFT, DWT, and NEW/GF yielded accuracy results of 85%, 85%, and 91% respectively when trained and tested using EEG snippets of data and validated using participant cross validation LOO on the 56 participants who watched Videos 1, 2a, and 2b (but excluding participants to self-identified or where EEG noise was excessive)

4.5 Transformation Method – Gradient Features

As described in the Methods section, the Gradient Features (GF) transformation method uses statistical measures to extract features from a series of coefficients, just as the DWT algorithms do. The difference between GF and DWT is that GF skips the band filtering and DWT transformation of the data prior to taking the statistical measures, and instead takes these same statistical measures of the raw EEG signal itself as well as on multiple levels of gradients.

The statistical measures selected for use in the final system arise for a series of experiments where different combinations of statistical measures were turned on and off and the accuracy results were compared. This testing was not completely rigorous in that not all 8,182 combinations of statistical measures were compared. Instead, different combinations of statistical measures were compared and when a good accuracy result that was comparable with DFT accuracy or better was achieved, comparisons were stopped. In this way, eight of the total possible thirteen compared statistical measures were used in the final system.

The statistical measures compared are shown in Table 13.

Once the statistical measures were selected, a comparison was made regarding the number of gradients needed to achieve a good accuracy result. The analysis was run using the statistical measures shown above, as well as the other settings found to be optimal from the section above and summarized in Table 7. The graph in Figure 34shows the accuracy achieved versus the different highest gradient level used. The first value of 88% uses only the eight statistical measures from the original EEG signal, or the 0^{th} level gradient, equivalently. The second value of 89% uses the eight values from gradient level 0, as well as the additional eight statistical measures taken from the first gradient. So the total number of features equals eight times the Highest Gradient Level Used plus one.

The conclusion from the above analysis shows that only the original signal plus its first and second derivative are needed to achieve a maximum accuracy measure. This is consistent with the cross disciplinary learning from control systems where the GF algorithm arises.

The final system therefore uses the GF algorithm up to level 2.

Equation Number in this paper	Name	Symbol
Equation 14	Entropy	$\boldsymbol{H(C)}$
Equation 16	Power	$\boldsymbol{P_C}$
Equation 17	Median	$\boldsymbol{Median_C}$
Equation 18	Minimum	$\boldsymbol{Min_C}$
Equation 19	Maximum	$\boldsymbol{Max_C}$
Equation 20	Slope	$\boldsymbol{\mu_{\nabla C}}$
Equation 21	Max over Mean	$\dfrac{\mathbf{max_C}}{\boldsymbol{\mu_C}}$
Equation 22	Number of Zero Crossings	$\boldsymbol{ZX_C}$

Table 13 Statistical Measures used in Final System

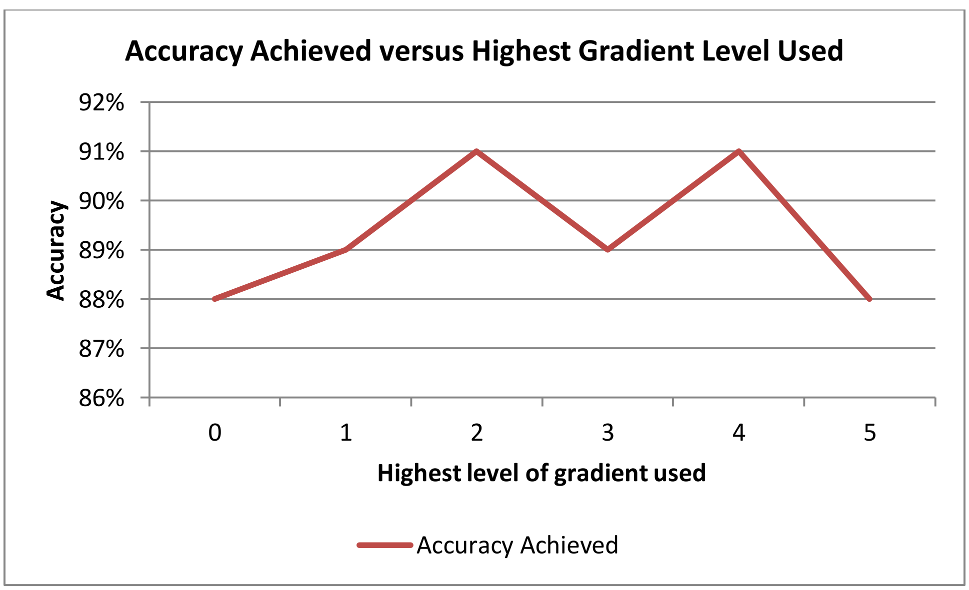


Figure 34 Graph comparing gradient levels

4.6 Noise Removal Threshold Selection Method

As described in the methods section, an optimal means of selecting a positive and negative threshold value is described, whereupon the researcher can use known information about the expected interval between noise events to select a threshold value. This method is proven to be a valid technique by analyses that compare accuracy of the system both when using and not using threshold based noise removal, and also by analyses that vary the values of the thresholds.

Threshold based noise removal is not so critical in the 40 participant experiments that only use Video1, but the algorithm was needed to get the highest accuracy when using all three videos in the data set (56 participants). This 56 participant accuracy was 91% using the GF transformation algorithm. The threshold based noise removal algorithm is therefore helpful in achieving maximum accuracy.

Figure 35 shows an example ten seconds of raw EEG data samples at 512 Hz, including a single possible ocular (blink) noise event. As can be seen in this example, the time between blink events in this sample is greater than one second (only one event occurs, and it is roughly in the center of the ten seconds of data). This is therefore an optimal candidate both for threshold based noise removal, as well as the method of threshold selection based on the expected time between noise occurrences.

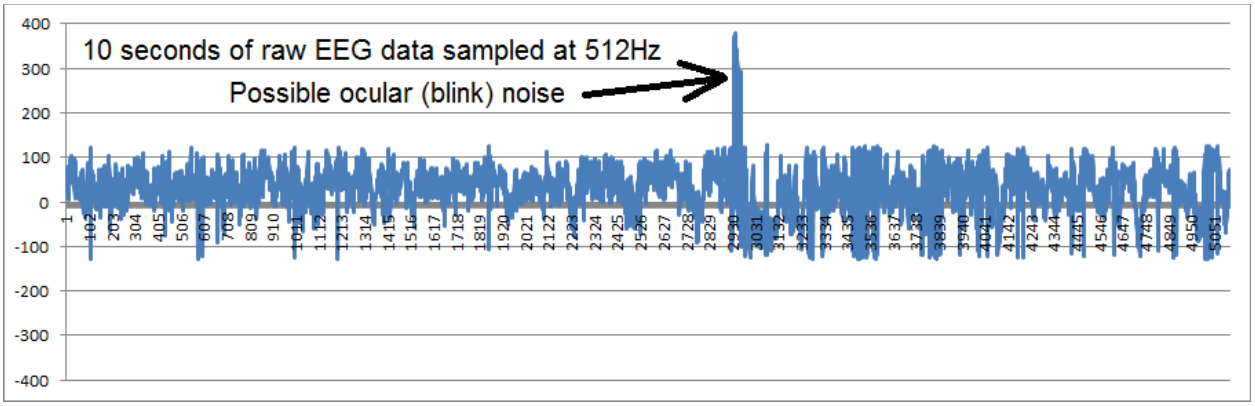


Figure 35 Ten seconds of raw EEG data showing ocular (blink) noise

The desired result is shown below in Figure 36, where the same ten seconds of EEG data is shown before and after threshold based noise removal. The noise removal technique shows a visual improvement of the EEG data with regard to eye blink noise removal, however it requires the selection of lower and upper bound thresholds to be selected. As discussed in the background and methods sections, the selection of these values has not been well defined until now, and have suffered from the issues of possible arbitrary manual selection, possible drift of equipment calibration, and the lack of connection between the threshold values and the desired goal of the noise removal algorithm – which is to remove noise events that occur "infrequently." To define the term infrequently means to select an expected time interval between these events. The method of threshold values traverses all of these issues. The results below confirm that indeed these values are optimally selected.

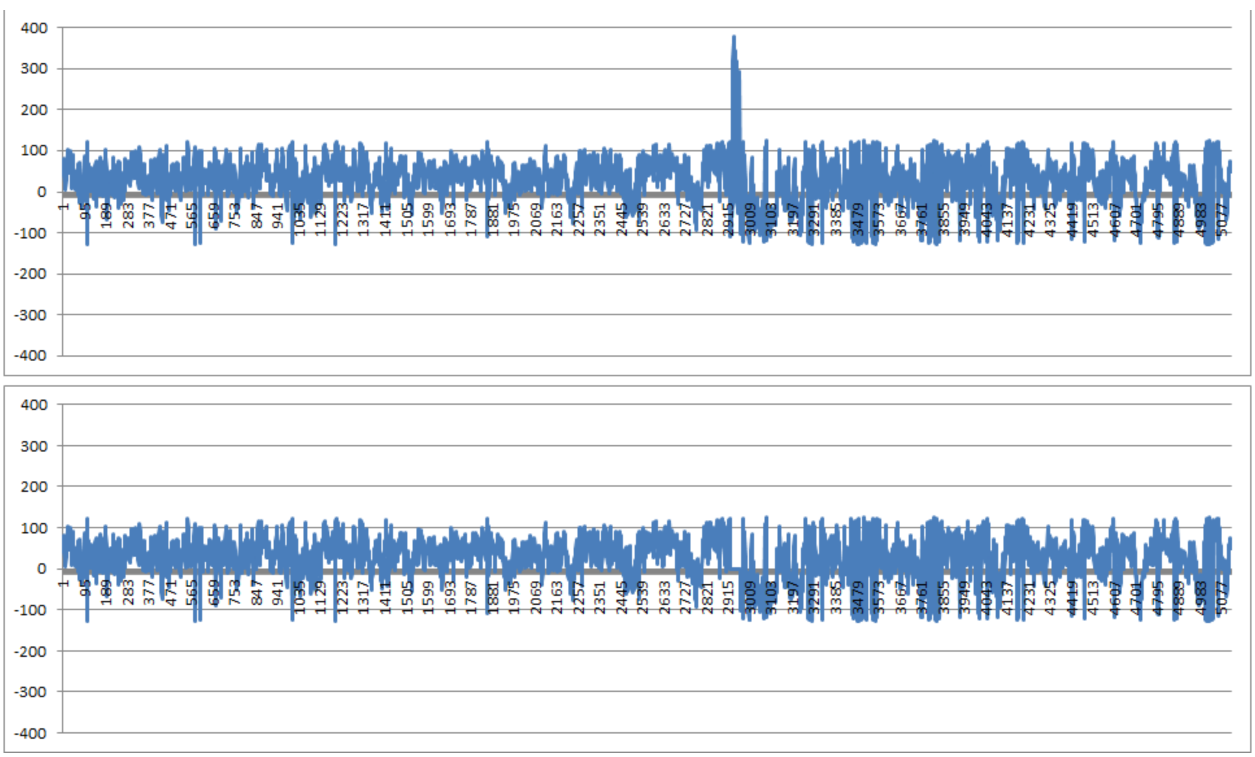


Figure 36 EEG data before and after threshold based noise removal

To prove that the threshold values themselves are optimal, one analysis compared expanding and contracting the range of the threshold values, and another analysis compared shifting the threshold values. The optimal threshold values were found using the novel visualization method based on expected intervals between noise occurrence, and these values were found to be -170 as the low threshold value, and 140 as the high threshold value (these are relative voltage integer levels as digitized by the EEG data collection hardware). Using these levels for the threshold based noise removal algorithm, and using the GF transformation algorithm, an accuracy measure of 91% was achieved. The distance between these values is (140) – (-170) = 310. The center, or average of these values is (140 -170) / 2 = -15

The first experiment changes the distance between the threshold values, keeping their center at -15. Figure 37 shows the range of different threshold distances that were used in this analysis. This helps visually show, that for at least this participant and these ten seconds of EEG data, there is no great surprise in the optimal threshold distance working the best in terms of accuracy.

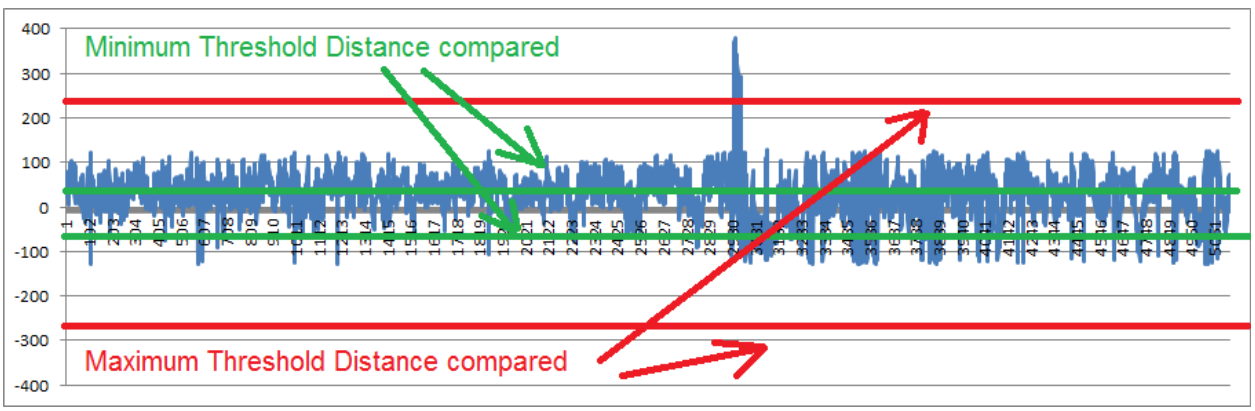


Figure 37 Different Threshold distances

As can be seen from Figure 38, the accuracy is at a maximum at the originally selected threshold values of -170 and 140 (the distance between them being 310).

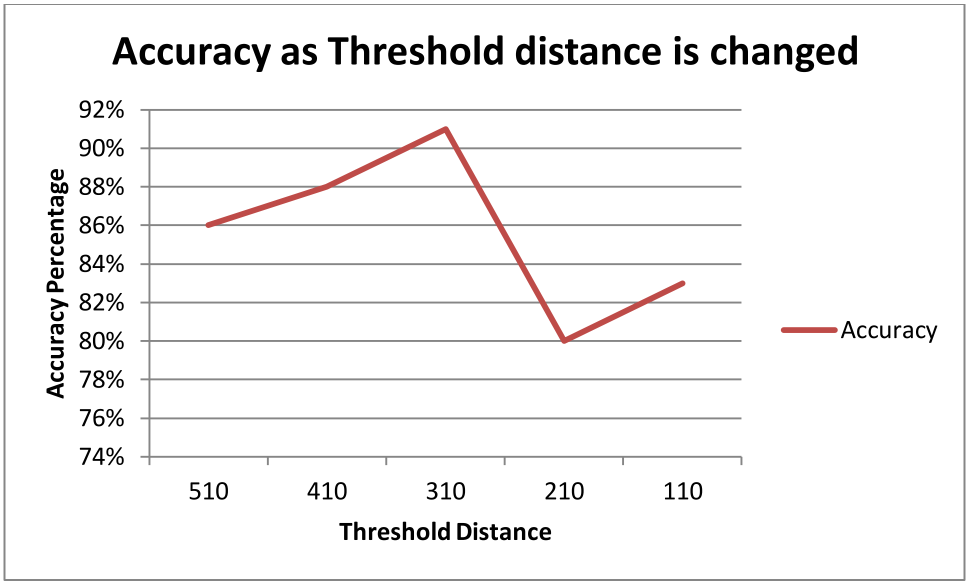


Figure 38 Effect of changing Threshold distance on accuracy

A second experiment was conducted to see if shifts in the center of the pair of threshold values impacted overall accuracy of the solution. In this analysis, the distance between the thresholds is kept the same (310) and instead the center (or average) of the threshold values are shifted. Figure 39 graphically depicts the range of threshold centers (in this case, shown by their minimum and maximum bounds) used in the above analysis. It is not visually obvious what the optimal threshold center (or distance) will be – even for this ten seconds of EEG data. Searching for the optimal thresholds across tens of thousands of seconds of EEG data would therefore prove an even more challenging endeavor.

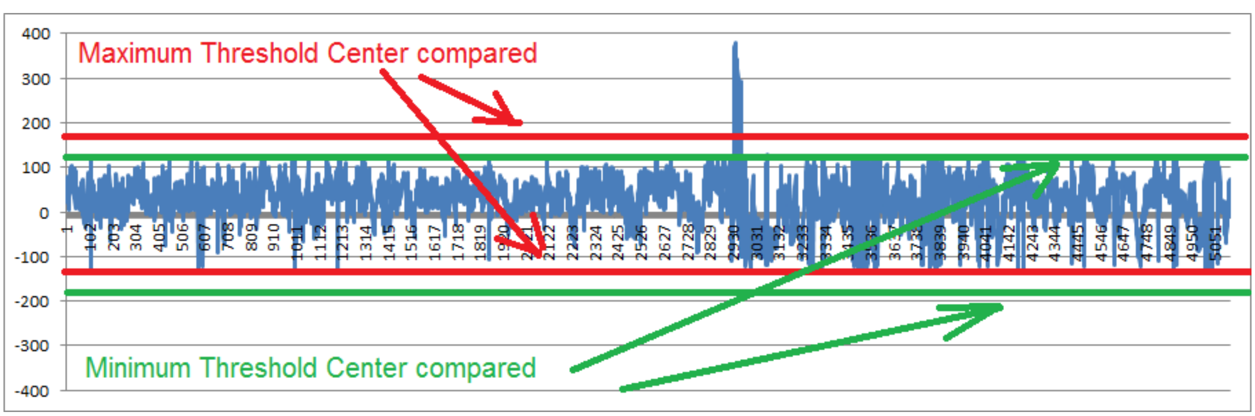


Figure 39 Different Threshold centers

As can be seen from Figure 40, the highest accuracy is achieved by using the threshold center selected by the novel algorithm and method. When the center of the thresholds is -15, the maximum accuracy is achieved, all other settings being equal.

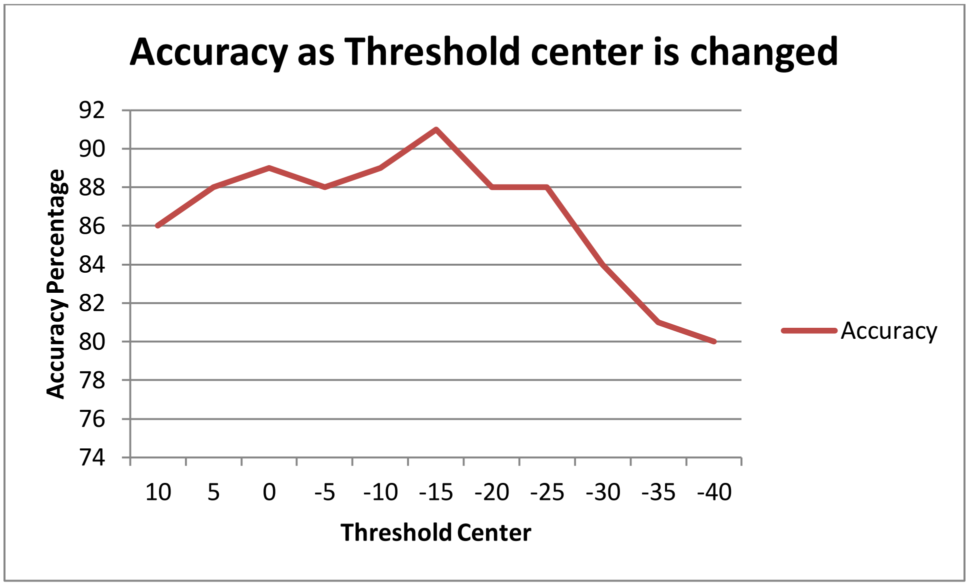


Figure 40 Effect of changing Threshold center on accuracy

4.7 Additional Analysis of Final System

The final system is actually three systems, based on the three different transformation algorithms DFT, DWT, and New GF. Analysis of the complete final system is both quantitative and qualitative. Quantitative results based on the 56 participants who each watched one of the three videos, and from whose EEG signals were extracted 84 seconds of data (half from the boring video segment and half from the attentiveness inducing segment of the video) were presented above. This section examines in more detail how the final system might be used by an operator on one or more participants in an effort to measure attentiveness towards short training videos.

For the additional quantitative analysis, the analysis uses all 69 participants. This means that the analysis is using 13 more participants who either self-identified with from the disability/disorder questions, or whose EEG data looked to be extremely contaminated with noise by visual inspection by the investigator. The reason the analysis is examining all 69 participants is because it may not be possible for the operator to ask participants to self-identify, and even this method is a limitation of the study, since none of the participants were asked to undergo a formal medical attention test. Also, it may not be possible for the untrained operator to look at EEG data to determine if it looks very contaminated with noise. Because of these two facts, the analysis is looking at the final system only if it was possible to collect the EEG data, and nothing else.

An accuracy measure was used based on 400 seconds of EEG data, 200 seconds of which were from the boring video segment, and 200 seconds were from the attentiveness inducing training video segment. After all of the pre-processing, automated noise removal, signal analysis, and feature extraction described above, each EEG epoch was presented to the machine learning algorithm. The number of false positives during the boring segment were counted for each participant and if it was less than 50% of that data then it was deemed to be a success. Similarly, the number of false negatives during the attentiveness segment were counted for that participant, and if the amounted to less than 50% of that data then it was also deemed to be a success. The overall success per participant was therefore evaluated at 0%, 50%, or 100% - and the total success percentage over all 69 participants was so calculated. The summary of the quantitative results are shown below in Table 14.

Qualitatively, analysis of the final system is possible from the point of view of examining the ensemble averaging across multiple participants to evaluate the overall video in terms of attentiveness inducing properties. This can help in identifying the most attentiveness inducing portions of the video.

Final System Type	Total Time to calculate Attentiveness Measure over 69x400=27,600 seconds of data	Total number of features per EEG epoch presented to the machine learning algorithm
DFT	46.3 seconds	69 (power for 1Hz bands from 2 to 80 Hz)
DWT	25.1 seconds	48 (8 statistical measures on 6 DWT coefficients for six EEG bands)
GF	8.1 seconds	24 (8 statistical measures on 0, 1st, and 2nd level gradients)

Table 14 Final System Quantitative Results

The most data collected was on Video1 that begins with a boring video segment of a candle burning and a title reading “relax and focus on your breathing” and ends with an attentiveness inducing training video having the participant fold paper. This analysis has 48 participants worth of data and averages the system’s attentiveness measure across the ensemble of participants to see a time varying track of how attentive the group of participants was for different sections of the video. This can be seen in Figure 41. This is a good image of how the final system will work, but it is a bad analysis of Video1 because it contains invalid data. Some portions of the data were actually used to train the system (the snippets) and so they must be removed from the image, and also from the smoothing algorithm, to have an accurate analysis of Video1. This is contained in Figure 42.

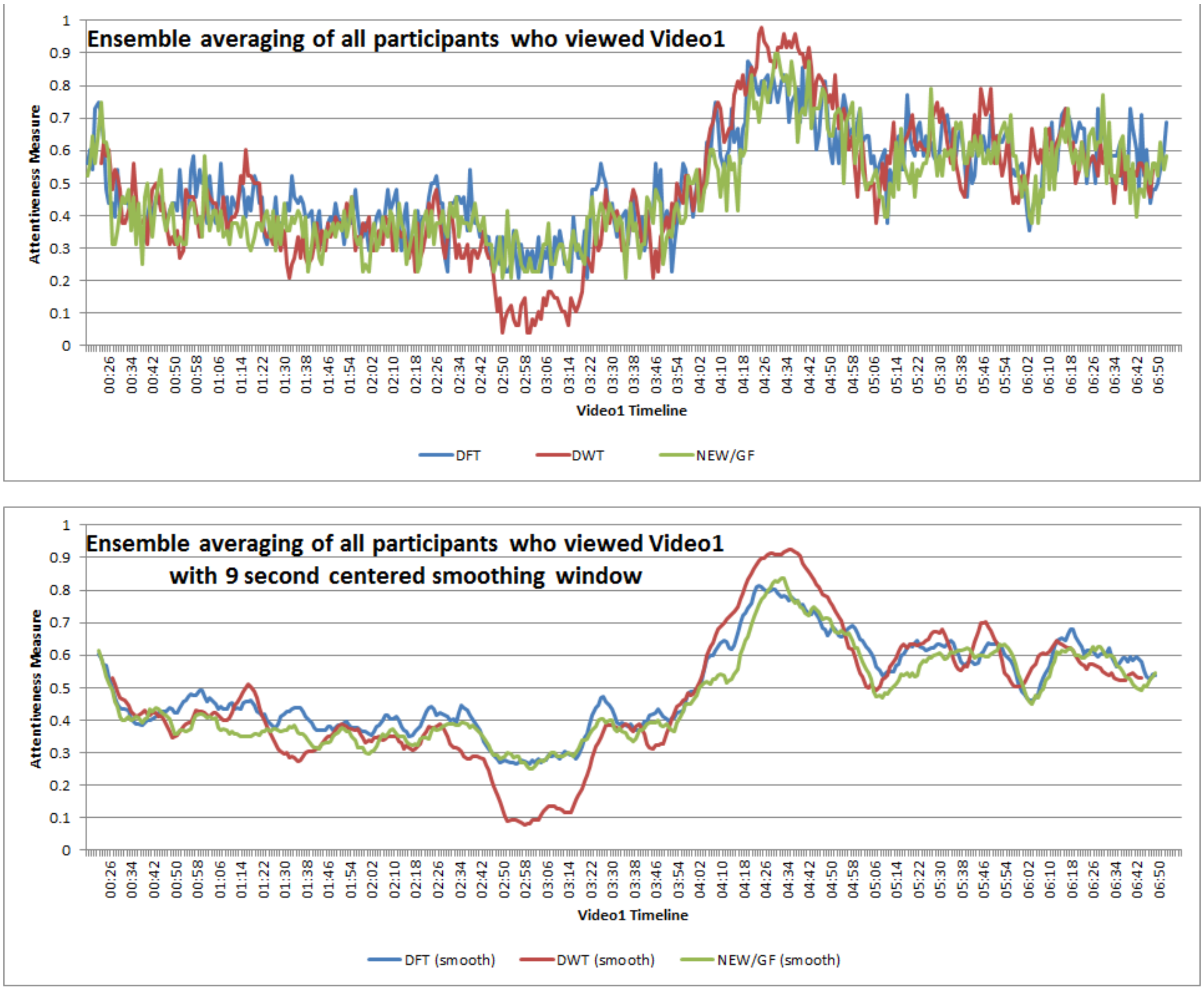


Figure 41 Ensemble average with invalid data

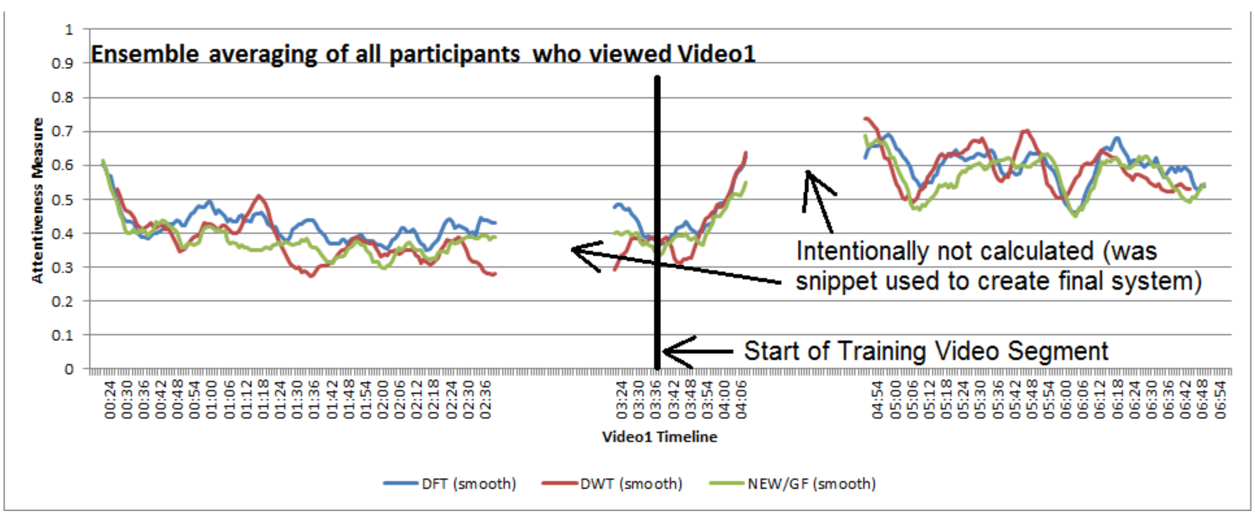


Figure 42 Ensemble averaging of Video1

No matter which of the three systems is used (DFT, DWT, or NEW/GF) Figure 42 clearly shows that participants are at first a little bit attentive even towards the boring candle burning video segment, and their attentiveness wanes over time. It can also be seen that it takes a few seconds to raise the brain into a state of attentiveness, even after the training video has begun. Perhaps most importantly to the operator of the system, areas can be seen within the training video that seem to capture more attentiveness on the part of the participants than other portions of the training video. Taking a closer look at that peak in attentiveness that seems to be taking place around 90 seconds into the training video (this is true no matter which of the three systems is used, and across all of the participants). Figure 43 shows a storyboard of the portion of Video1 where attentiveness measurements are beginning to peak. This portion of the training video is the second of several times where the instructor has finished describing what the participant is being asked to do, and the camera fades to a segment where the instructor is modeling what the participant is being asked to do.

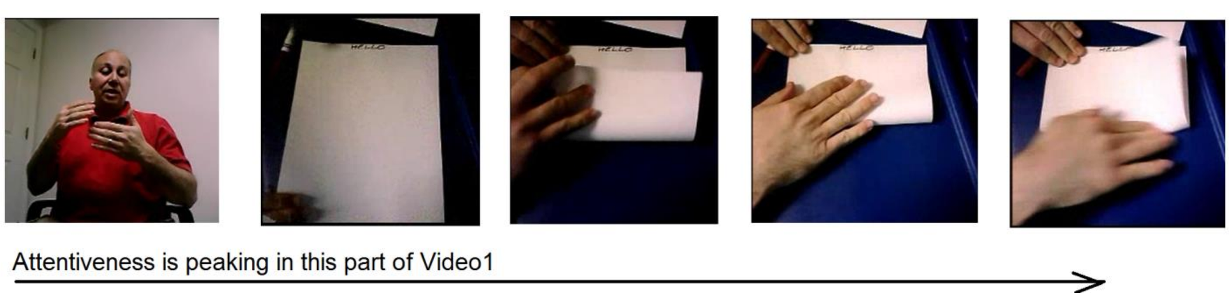


Figure 43 Attentiveness Peak in Video1 (Storyboard)

Modeling is a very important part of education, and it is where the instructor goes through the very motions that the student (participant) is going to have to perform. The short training video in this research was purposely designed to include modeling because it is known that this was an attentiveness inducing part of a training lesson. It is qualitative confirmation that the system is performing as requested, using EEG to measure attentiveness towards short training videos. As mentioned above, this peak of attentiveness at around 90 seconds into the training video is the second time when modeling takes place. The first time is around 50 seconds into the training video, and it is where the student sees the instructor's hand writing the word "Hello" on the paper. Other modeling events also take place in the video, including the second fold of the paper, and putting the folded paper into the envelope.

The graph of attentiveness is very useful, but it leaves the operator examining a chart, and then switching back to the video and scrolling back and forth to create the above storyboard in synch with the chart. This is time consuming and laborious. It might be possible to simultaneously show the chart values (perhaps as a meter needle or colored spot) and the video playing back dynamically, which could be a future improvement to the system. Even with the simultaneous dynamic display of the chart and video playback, the operator is left to glance back and forth from the video to the attentiveness meter. An even better solution is to replace the video audio soundtrack with an audio sound that represents the attentiveness of a participant or an ensemble of participants. This innovation has had some beginning development as part of this research and is described in the methods section and here in the results section.

Similar ensemble averages exist for the other videos, although for Video2a and Video2b there were fewer participants (10 and 11 respectively) and so the graphs do not benefit from the larger number of students with which to average. These are shown in Figure 44 and Figure 45, respectively. As described earlier, Video2a started with a boring lecture (instead of a candle). Judging from the diagram below, this turned out to be even more boring than the candle. It took participants watching Vide2a longer to enter into the attentive state. This may be caused in part due to the fact that the same lecturer gave the boring lecture, and then the attentiveness inducing short training video. Also, for the training segment, a different range was used for the snippet, and so the first modeling portion (writing the word "Hello") is now seen. Some of these effects mentioned about Video2a may also be due to the fact that both Video2a and Video2b had many fewer participants, and as such do not have the benefit of a larger ensemble.

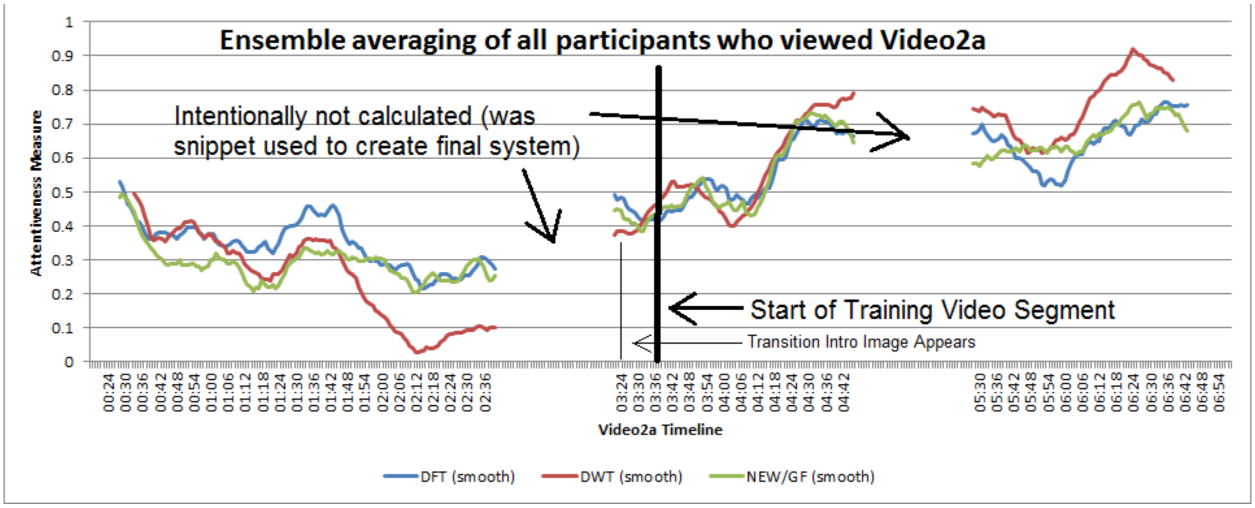


Figure 44 Ensemble Average of Video2a

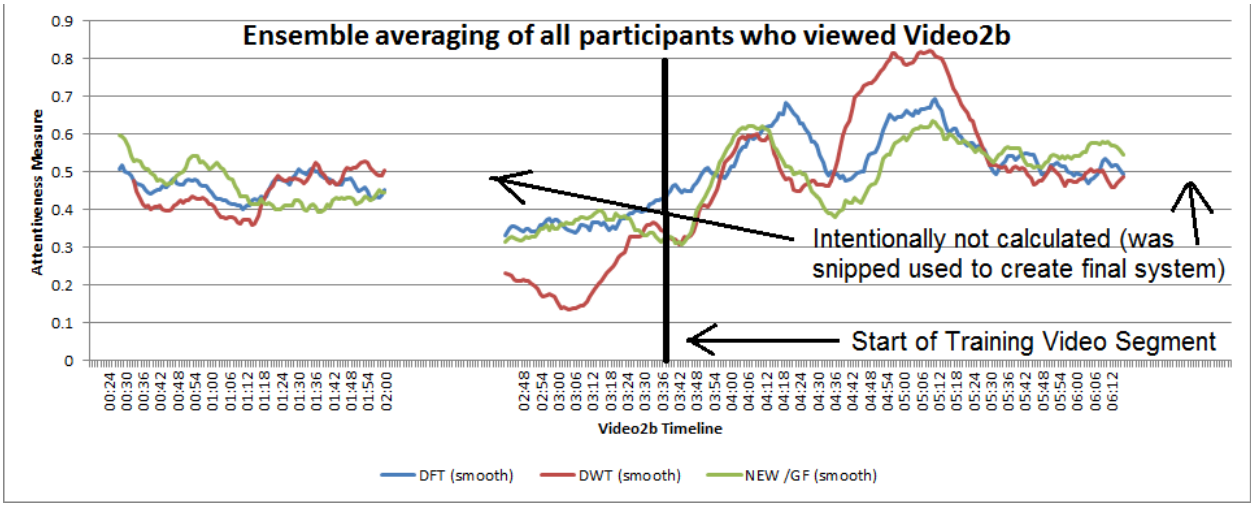


Figure 45 Ensemble Average of Video2b

4.8 Pseudo Audio Method for Operator Analysis

As mentioned in the Methods and Results sections above – it would be desirable if an operator using the system could listen to the EEG and/or the internal representation of the EEG signal (in the form of attentiveness features). This is achieved as described in the methods section, and a multimedia file accompanies this paper to demonstrate this.

4.9 Triangulation with Qualitative Data

The system also can produce an attentiveness graph for an individual participant. Figure 46 shows such an output of the GF final system which is passed through a Centered 40 second averaging to smooth results of a participant watching Video1, with the transition from the candle video and the attentiveness inducing training video taking place midway through the video.

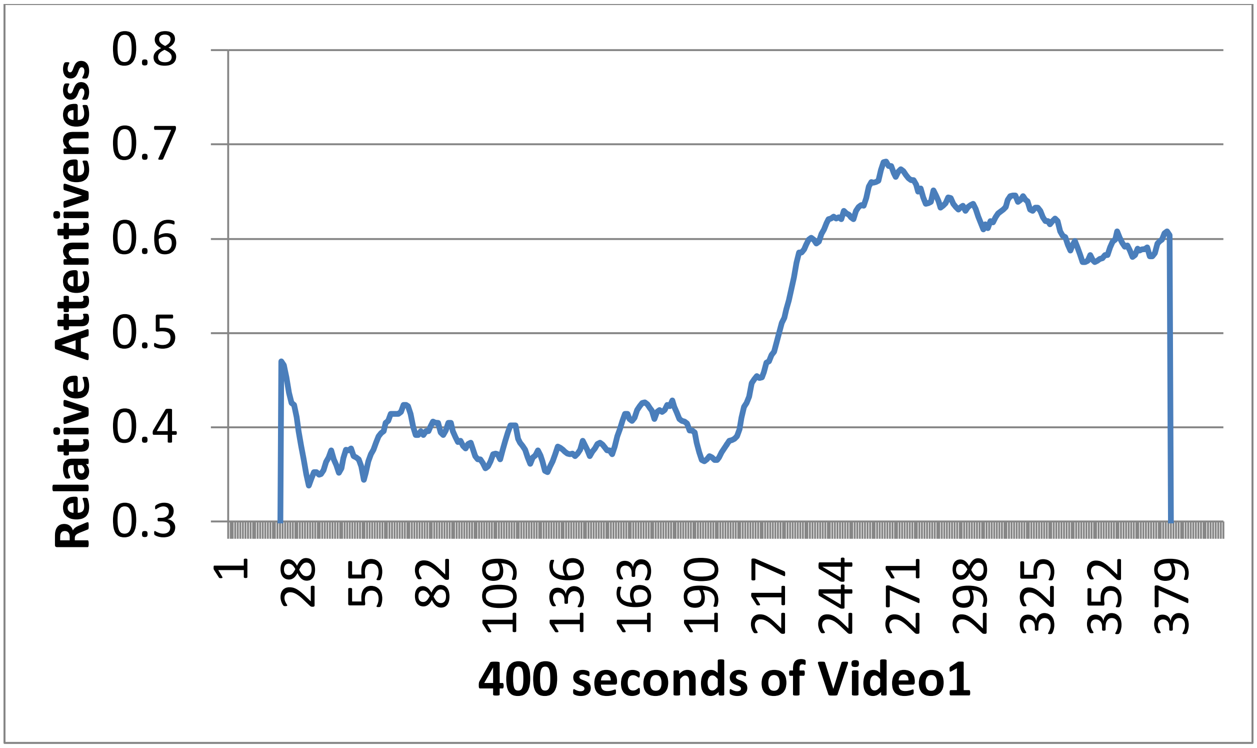


Figure 46 Typical Attentiveness Graph – Participant 2

In the above, we can see the typical initial attentiveness that begins the boring half of the video, and this attentiveness quickly wanes. We also see the typical ramp up of attentiveness, as the participant takes some time to come out of their state of boredom. Also typically, we can see that the training video half does not keep the attentiveness at a peak level continuously, especially as participants become accustomed to getting instructions and following them.

Additional examples of these sorts of graphs, as well as anomalies triangulated with qualitative data, can be found in Appendix C.

4.9.1 Limitations of the Study

There are some limitations of the study with regard to the participant self-reporting. By using self-reported LD and ADHD status versus having formal attention testing, medically labeled individuals, it is not possible to triangulate effectively based on this single questionnaire indicator. Indeed, the accuracy of the final system is demonstrated using both toe old and new questionnaires PreQA and PreQB. In addition, the study does not investigate the variety of developmental aspects and training aspects. A more rigorous investigation focused on these areas would be needed to expand this research, which currently largely focused on signal analysis of the EEG data.

4.9.2 Triangulation Results Summary

In summary, triangulation of quantitative attentiveness measures, our expectation of attentiveness towards short training videos (and boring videos) and qualitative events and self-described feelings and thoughts of participants, proved a useful reinforcement of the success of the research and the final system. Again and again, both in attentiveness and inattentiveness cases, the qualitative data are found to triangulate with the quantitative measure of attentiveness by the DFT System.

4.10 Mother Wavelet and Decomposition Level Selection Method

As described in the Methods section, an algorithm is implemented in this research where both the mother wavelet shape and the decomposition level are selected based on the existing EEG data and the goal of finding a mother wavelet that most closely approximates the important features in the EEG signal at that specific decomposition level (at that scale, equivalently). Since the wavelet is essentially correlated with the EEG signal, the search for the maximum value is proposed.

Ideally, it would be best to confirm these proposed optimally selected wavelet shapes and decomposition levels by comparing all possibilities in the DWT system itself, and see if they match with the highest accuracy level. The problem with this is that the accuracy level graphed against all possible combinations of wavelet shape and decomposition involved interaction between one wavelet shape for the Alpha band against another wavelet shape for the Beta band, etc.

In an effort to remove this obstacle, an analysis was conducted with the DWT system using only the Beta band. This lowers the total accuracy of the system, but relative accuracies can be compared when the Beta band is analyzed using one wavelet shape or level against another. Figure 47 shows the results described in the Methods section for the Beta band on the left, and the results of the relative accuracy of the DWT system when using only the Beta band. As expected, the accuracy levels on the right are lower than the 89% achieved when multiple bands are used, but one might want to look only at the relative accuracy levels. Although the graph on the right is definitely noisier, it is clear that the accuracy goes down as the decomposition level rises (moves towards the back of the graph). Other conclusions (such as selection of the specific wavelet shape) are harder to discern, and further research in this area may be warranted.

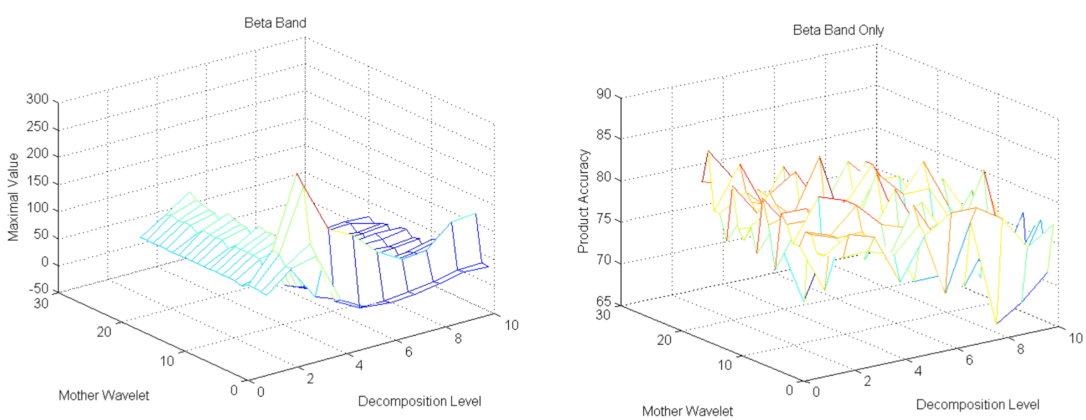


Figure 47 Beta band wavelet shape and decomposition level maximal (left) and accuracy (right) comparison

4.11 Comparison of Machine Learning Algorithms

An extensive comparison of machine learning algorithms was conducted in the previously published pilot study (Nussbaum and Hobson-Hargraves 2013). The conclusion at that time was the use of kNN was optimal for this application.

Nevertheless a comparison of machine learning algorithms on the current EEG data set is worth performing, if only to re-assess whether or not kNN continues to be an optimal selection of machine learning algorithm. The comparison is performed on the DFT, DWT, and GF Systems, except the machine learning algorithm (Baseline kNN) is replaced with non-baseline machine learning algorithms using 10-fold cross validation (using data from all participants).

The machine learning algorithms seem to do best with the DWT system, but in all cases, the kNN algorithm continues to perform best, reconfirming the decision to choose this algorithm as the optimal one for this application. A diagram of these results is shown in Figure 48, Figure 49, and Figure 50.

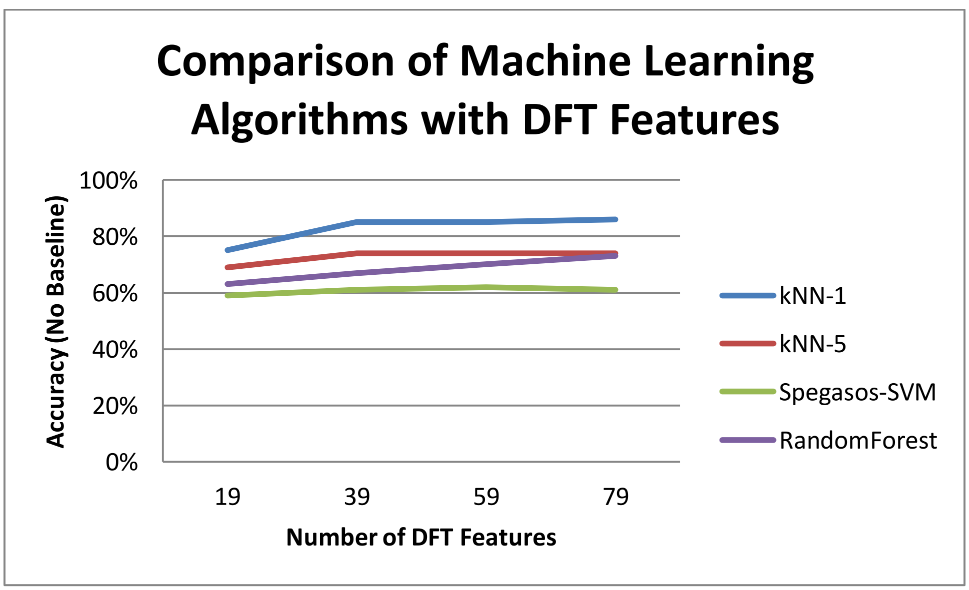


Figure 48 Comparison of Machine Learning Algorithms with DFT Features

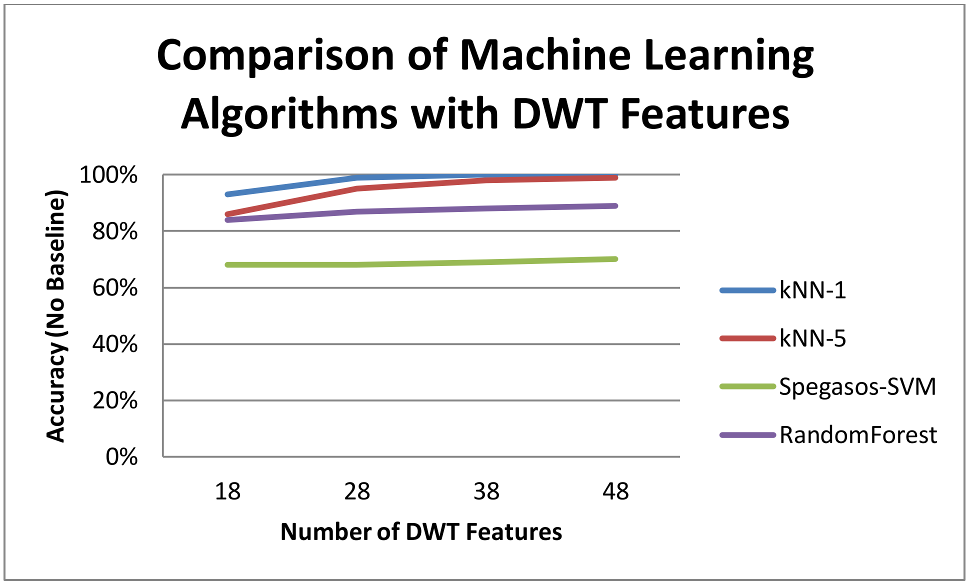


Figure 49 Comparison of Machine Learning Algorithms with DWT Features

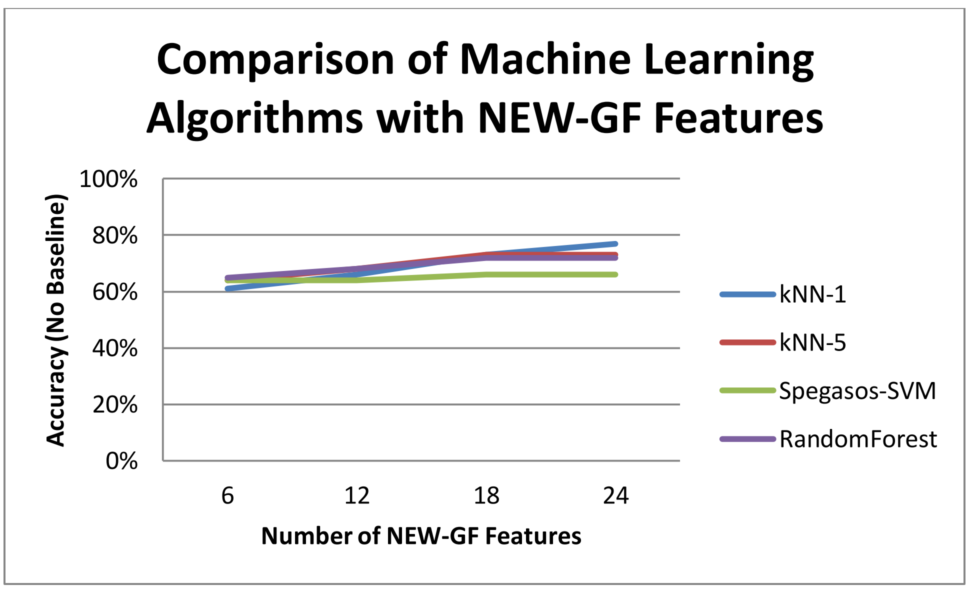


Figure 50 Comparison of Machine Learning Algorithms with NEW-GF Features

CHAPTER 5 Conclusions

The research has created three systems, each of which is capable of determining attentiveness indicators from participants watching short training videos by using only a single dry electrode EEG device wirelessly connected to a laptop computer. The performance of all three systems is very good, having been tested against a large number of participants, as well as different boring video and attentiveness inducing short training video examples. The systems were then tested against a larger amount of data, and although the results are not identical, they are sufficiently similar when triangulated with qualitative data to confirm they are actually measuring indicators of attentiveness. The three systems are each based on different signal analysis and feature extraction methods (DFT, DWT, and a new GF signal analysis method), and the fact that they arrive at largely the same accuracy level and also largely identify which seconds of a short training video generate the most attentiveness, help to confirm that the research has been successful in developing these systems.

A number of novel methods and algorithms were developed as part of this research, and these also performed well and helped to create the successful end results.

CHAPTER 6 Future Work

The system already has the potential to contribute to future improvement of short training videos. Through modern biometric electrical engineering, the research has confirmed that modeling, especially with regard to the videotaping of the instructors hands performing a task that the student will have to perform, does indeed increate attentiveness in users watching these videos. This suggests additional future work to expand the variety of short training videos and related educational methods can be tested with the new methods and apparatus.

Improvements to the actual system itself will involve making the operator interface more user friendly, including an automated excessive EEG noise detection method, and packaging the product with a software environment that facilitates use alongside short training videos from a variety of sources.

Another opportunity is the examination of results when the novel technique of EEG pseudo-audio generation is used for other EEG applications (such as sleep study and epilepsy seizure). Separate from this paper, multimedia files are available so the reader can see and hear the novel method as it would be used in a deployed setting.

Finally, many of the algorithms seem to be extensible to other EEG applications. The idea of associating measured attentiveness with learning disabilities and attention deficit disabilities, perhaps extending the use to the research regarding different types of autism, may be of interest.

CHAPTER 7 Bibliography

Bakera, R. D., S. D'Mellob, M. Rodrigo, and A. Graesser. "Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments." *Int. J. Human-Computer Studies* 68 (2010): 223–241.

Belle, Ashwin, Rosalyn Hobson Hargraves, and Kayvan Najarian. "An Automated Optimal Engagement and Attention Detection System Using Electrocardiogram." *Computational and Mathematical Methods in Medicine* 2012, no. Article ID 528781 (2012): 1-12.

Berka, C., et al. "Real-Time Analysis of EEG Indexes of Alertness, Cognition, and memory Acquired with a Wireless EEG Headset." *International Journal of Human-Computer Interaction* 17 (2) (2004): 151-170.

BIOPAC Systems, Inc. "MP Systems hardware Guide." PDF File 07.22.11 (2011).

Chakraborty, A., P. Bhowmik, S. Das, A. Halder, A. Konar, and A. K. Nagar. "Correlation between Stimulated Emotion Extracted from EEG and its Manifestation on Facial Expression." *Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics* 2009 (2009): 3132-3137.

Chalfoun, Pierre, Soumaya Chaffar, and Claude Frasson. "Predicting the Emotional Reaction of the Learner with a Machine Learning Technique." *Workshop on Motivaional and Affective Issues in ITS,* ITS'06 (2006).

Crowley, K., A. Sliney, I. Pitt, and D. Murphy. "Evaluating a Brain-Computer Interface to Categorise Human Emotional Response." *2010 10th IEEE International Conference on Advanced Learning Technologies*, 2010: 276-278.

de Munck, Jan, Petra van Houdt, Sonia Goncalves, Erwin van Wegen, and Pauly Ossenblok. "Novel artefact removal algorithms for co-registered EEG/fMRI based on selective averaging and subtraction." Edited by Elsevier. *NeuroImage* 64 (2013): 407-415.

De Vico Fallani, F., and et al. "Structure of the cortical networks during successful memory encoding in TV commercials." *Clinical Neurophysiology* 119 (2008): 2231-2237.

D'Mello, Sidney K., Scotty D. Craig, and Art C. Graesser. "Multimethod assessment of affective experience and expression during deep learning." *Int. J. Learning Technology* Vol. 4, no. Nos. 3/4 (2009): 165-187.

D'Mello, Sydney, and Art Graesser. "Dynamics of affective states during complex learning." *Learning and Instruction* (Elsevier) 22 (2012): 145-157.

Epstein, Charles. "The Nuts and Bolts of Electroencephalography." *Sleep Med Clin* (Elsevier) 7 (2012): 1-12.

Frasson, Claude, and Pierre Chalfoun. "Managing Learner's Affective States in Intelligent Tutoring Systems." Edited by R. Nkambou et al. *Advances in Intelligent Tutoring Systems* (Springer-Verlag Berlin Heidelberg) SCI, no. 308 (2010): 339–358.

Freire, Luciana, Arezes Pedro Miguel Lopes, and Jose Creissac Campos. "A literature review about usability evaluation methods for e-learning platforms." *IOS Press*, 2012: 1038-1044.

Gargiulo, Gaetano, et al. "A new EEG recording system for passive dry electrodes." Edited by Elsevier. *Clinical Neurophysiology* 121 (2010): 686-693.

Goldfine, Andrew, Jonathan Victor, Mary Conte, Jonathan Bardin, and Nicholas Schiff. "Determination of awareness in patients with severe brain injury using EEG power spectral analysis." Edited by Elsevier. *Clinical Neurophysiology* 122 (2011): 2157-2168.

Green, T.M., and K. Najarian. "Correlations between emotion regulation, learning performance, and cortical activity." *Proceedings of the 29th Annual Conference of the Cognitice Science Society.* Nashville, TN, 2007. 1764.

Groullier, Frederic, Laurent Vercueil, Alexander Krainik, Christoph Segebarth, Phillippe Kahane, and Oliver David. "A comparative study of different artefact removal algorithms for EEG signals acquired during functional MRI." Edited by Elsevier. *NeuroImage* 38 (2007): 124-137.

Gwin, Joseph, Klaus Gramann, Scott Makeig, and Daniel Ferris. "Removal of Movement Artifact From High-Density EEG Recorded During Walking and Running." *J Neurophysiol* (The American Physiological Society) 103 (2010): 3526-3534.

Hayashi, Katsuki, Keitarou Ishihara, Haruhiko Hashimoto, and Koji Oguri. "Individualized Drowsiness Detection during Driving by Pulse Wave Analysis with Neural." *Intelligent Transportation Systems, 2005. Conference Proceedings.*, 2005: 901-906.

Heller, Rachelle S., Cheryl Beil, Kim Dam, and Belinda Haerum. "Student and Faculty Perceptions of Engagement in Engineering." *Journal of Engineering Education*, July 2010: 253-261.

Heraz, Alicia, and Claude Frasson. "Predicting the Three Major Dimensions of the Learner's Emotions from Brainwaves." *World Academy of Science, Engineering and Technology* 31 (2007): 323-329.

Hobson-Hargaves, Rosalyn. "Class Notes from VCU EGRE 630 "Neural Networks", ." 2011.

Ito, Shin-ichi, Yasue Mitsukura, Katsuya Sato, Shoichiro Fujisawa, and Minoru Fukumi. "Study on Association between User's Personality and Individual Characteristic of Left Prefrontal Pole EEG Activity." *2010 Sixth International Conference on Natural Computation (ICNC 2010)*, 2010: 2163-2166.

Jung, T., S. Makeig, M. Stensmo, and T. Sejnowski. "Estimating Alertness from the EEG power Spectrum." *IEEE Transactions on bio-medical engineering* 44, no. 11 (1997): 60-70.

Kay, R. H. "Exploring the use of video podcasts in education: A comprehensive review of the literature." *Computers in Human Behavior* 28 (2012): 820–831.

Khare, Vijay, Jayashree Santhosh, Sneh Anand, and Manvir Bhatia. "Classification of Five mental Tasks from EEG Data Using Neural Networks Based on Principal Component Analysis." *The IUP Journal of Science & Technology* 5, no. 4 (2009): 31-38.

Khosrowabadi, Reza, Michel Heijnen, Abdul Wahab, and Hiok Chai Quek. "The Dynamic Emotion Recognition System Based on Functional Connectivity of Brain Regions." *2010 IEEE Intelligent Vehicles Symposium*, 2010: 377-381.

Kort, Barry, Rob Reilly, and Rosalind Picard. "An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy—Building a Learning Companion." *Proc. Int. Conf. Advanced Learning Technologies* 2002 (2002): 43–48.

Lee, Seung Hyun, Berdakh Abibullaev, Wong-Seok Kang, Yunhee Shin, and Jinung An. "Analysis of Attention Deficit Hyperactivity Disorder in EEG Using Wavelet Transform and Self Organizing Maps." *International Conference on Control, Automation and Systems 2010.* Gyeonggi-do, Korea: ICROS, 2010. 2439-2442.

Lee, Seung Hyun, Berdakh Abibullaev, Won-Seok Kang, Yunhee Shin, and Jinung An. "Analysis of Attention Deficit Hyperactivity Disorder in EEG Using Wavelet Transform and Self Organizing Maps." *International Conference on Control, Automation and Systems 2010 in KINTEX*, 2010: 2439-2442.

Lester, James C., Stuart G. Towns, and Patrick J. Fitzgerald. "Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents." *International Journal of Artificial Intelligence in Education* (IJAIED) 10 (1999): 278-291.

Li, Qi, and Donald W. Tufts. "Principal Feature Classification." *IEEE Trans.* 8, no. 1 (1997): 155-160.

Lisetti, C., and F. Nasoz. "Using Noninvasive Wearable Computers to Recognize Human ." *EURASIP Journal on Applied Signal Processing* 2004, no. 11 (2004): 1672-1687.

Ma, Junshui, Peining Tao, Seyvinc Bayram, and Vladamir Svetnik. "Muscle artifacts in multichannel EEG: Characteristics and reduction." *Clinical Neurophysiology* (Elsevier) 123 (2012): 1676-1686.

MacLean, Mary, Karen Arnell, and Kimberly Cote. "Resting EEG in alpha and beta bands predicts individual differences in attentional blink magnitude." Edited by Elsevier. *Brain and Cognition* 78 (2012): 218-229.

Mostow, J., K. Chang, and J. Nelson. "Toward Exploiting EEG Input in a Reading Tutor, Project LISTEN, School of Computer Science, RI-NSH 4103." *Artificial Intelligence in Education: 15th International Conference, AIED 2011*, 2011: 230-237.

Murugappan, M. "Human Emotion Classification using Wavelet Transform and KNN." *2011 International Conference on Pattern Analysis and Intelligent Robotic*, 2011: 148-153.

Nagashino, H., et al. "Application of Neural Networks to Dynamics Identification by EEG." *Seventh International Conference on control, Automation, Robotics And Vision* 1 (2002): 554-559.

Najarian, Kayvan, and Robert Splinter. *Biomedical Image and Signal Processing.* Second Edition. Boca Raton, FL: CRC Press, 2012.

NeuroSky Co. "MindWave Mobile: User Guide." Feb 11, 2012.

Nussbaum, Paul Alton, and Rosalyn Hobson-Hargraves. "Pilot Study-The Use of Electroencephalogram to Measure Attentiveness towards Short Training Videos." *International Journal of Advanced Computer Science and Applications* 4, no. 3 (2013): 176-182.

Palaniappan, Ramaswamy. "Utilizing Gamma Band to Improve Mental Task Based Brain-Computer Interface Design." *Neural Systems and Rehabilitation Engineering, IEEE Transactions on Neural Systems and Rehabilitation Engineering* 14, no. 3 (2006): 299-303.

Pasqualotto, E., S. Federici, and M. Belardinelli. "Toward functioning and usable BCI." *Disability and Rehabilitation:* 7 (2) (2012): 89-103.

Rebolledo-Mendez, G., et al. "Assessing NeuroSky's Usability to Detect Attention Levels in an Assessment Exercise." *Human-Computer Interaction, Part I, HCII 2009* LNCS 5610 (2009): 149-158.

Renvoise, P., and C. Morin. *Neuromarketing.* SalesBrain LLC, 2007.

Rumelhart, D., B. Widrow, and M. Lehr. "The Basic Ideas in Neural Networks." *Communications of the ACM* 37, no. 3 (1994): 86-92.

Safieddine, Doha, et al. "Removal of muscle artifact from EEG data: comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches." *EURASIP Journal on Advances in Signal Processing* (Springer) 127 (2012): 1-15.

Sandberg, David, Torbjörn Åkerstedt, Anna Anund, Göran Kecklund, and Mattias Wahde. "Detecting Driver Sleepiness Using Optimized Nonlinear Combinations of ." *IEEE trans. Intelligent transportations systems* 12, no. 1 (2010): 97-108.

Selvan, S., and R. Srinivasan. "Removal of Ocular Artifacts from EEG Using an Efficient Neural Network Based Adaptive Filtering Technique." *IEEE Signal Processing Letters* 6, no. 12 (1999): 330-332.

Subasi, Abdulhamit, M. Kemal Kiymik, Mehmet Akin, and Osman Erogul. "Automatic recognition of vigilance state by using a wavelet-based artificial neural networks." *Neural Computations and Applications* 14, no. 1 (2005): 45-55.

Uusberg, Andero, Helen Uibo, Kairi Kreegipuu, and Juri Alli. "EEG alpha and cortical inhibition in affective attention." *International Journal of Psychophysiology* (Elsevier), 2013: 1-11.

Watson, David, Lee Anna Clark, and Auke Tellegen. "Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales." *Journal of Personality and Social Psychology* Vol. 54, no. No. 6 (1988): 1063-1070.

Welling, Max. "Fisher Linear Discriminant Analysis." *University of Toroto Computer Science Department – classroom notes*, 2012.

Wilson, Beth J., and Thomas D. Bracewell. "Alertness Monitoring Using Neural Networks for EEG Analysis." *Neural Networks for Signal Processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop* 2 (2002): 814-820.

Appendix A Tables Comparing Prior Art

Ref.	Application / Experiment	Position of EEG sensors
(Khare, et al. 2009)	Classify mental state such as being asked to imagine moving their right hand, or doing simple math problems.	Eight standard positions C3, C4, P3, P4, O1 O2, F3 and F4
(Chakraborty, et al. 2009)	Participants watch emotion eliciting videos, and the EEG is correlated to their facial expression.	One position, either F3 or F4
(De Vico Fallani and al 2008)	Users watch a long documentary interspersed with commercials, and are asked days later which they remember.	Extending the International 10-20 Standard, 64 electrodes are used in addition to earlobe ground for reference
(Selvan and Srinivasan 1999)	This experiment seeks to identify EEG signals arising from ocular artifacts such as blinking, rolling eyes.	Positions F7 – T3 – also additional non-standard electrodes around the eyes
(Jung, et al. 1997)	Measure of alertness towards Navy sonar monitoring tasks	two midline sites, one central (Cz) and the other midway between parietal and occipital sites (Pz/Oz)
(Nagashino, et al. 2002)	Prediction of future EEG signals from past signals, participant has eyes open or closed.	All of the electrode locations of the International 10-20 system are used, however one at a time.
(Subasi, et al. 2005)	EEG to categorize vigilance state, versus drowsy or sleepy.	A single C3 electrode with an A2 earlobe referential
(Palaniappan 2006)	Use EEG to determine mental activity of participant, such as imagining visual counting, multiplication, etc.	Positions C3, C4, P3, P4, O1, and O2.
(Ito, et al. 2010)	EEG used to match listeners liking or not liking music, to their personality profile from 50 question exam	Single electrode FP1 and earlobe referential
(Khosrowabadi, et al. 2010)	Pictures and music attempt to elicit emotions, confirmed by questionnaire. EEG is them correlated to emotion	EEG is measured at points F3, F4, T7, C3, CZ, C4, T8, P3, and P4
(S. Lee, B. Abibullaev and W.-S. Kang, et al. 2010)	EEG differentiates between ADHD and mental retardation in children ages 10 – 13	Two frontal lobe locations
(Murugappan 2011)	EEG predicts emotion elicited from watching a video.	64 electrodes, an extension of the 10-20 system
(Rebolledo-Mendez, et al. 2009)	Measure how attentive students are while taking an exam	Single dry electrode (Fp2) and A2 earlobe referential.
(Crowley, et al. 2010)	Measure stress as users are asked to solve puzzles (tower of Hanoi) and tasks (Stroop Test) faster and faster.	Single dry electrode (Fp2) and A2 earlobe referential.
(Mostow, Chang and Nelson 2011)	EEG identifies if children and adults have an easy or a hard reading task.	Single dry electrode (Fp2) and A2 earlobe referential.
(Berka, et al. 2004)	Alertness measure as participants simulate defending against incoming enemy planes. Development of the B-Alert system (now supplied as part of the Biopac EEG system)	Bipolar sensors (one left and one right) at the positions of Fz, Cz, POz.
(Gargiulo, et al. 2010)	Comparing wet and dry EEG electrodes	Dry electrode signal almost identical to wet. Correlation coefficient between signal from dry and average signal from 4 surrounding wet electrodes is > 0.85

Table 15 Survey of EEG Sensor Positioning Methods selected by researchers

Reference	Noise Decontamination Methods
(Berka, et al. 2004)	B-Alert identifies and decontaminates fast and slow eye blinks, and ID's and reject data contaminated with EMG, amp. saturation, or excursions due to motion artifacts.
(Belle, Hobson Hargraves and Najarian 2012)	First a band pass filter is applied, and then individual band pass filtering is applied for each of the EEG bands under examination.
(Chakraborty, et al. 2009)	Peak and average power for the bands (frequency) and (time domain) are used. 16 Kalman filter coefficients, and (spatio-temporal) 132 wavelet coefficients are compared. Data across 5 emotions, eliminated correlation by PCA.
(Crowley, et al. 2010)	Only as provided by NeuroSky headset (Neurosky providing "Meditation" value).
(De Vico Fallani and al 2008)	Trained neurologists visually inspected the acquired EEG data and rejected EEG epochs containing muscle or gross EEG artifacts. Subsequently, the EEG signals were filtered from the residual EOG: electrooculogram activity with standard ordinary least squares procedure, baseline-adjusted and low-pass filtered at 45 Hz
(Gargiulo, et al. 2010)	Standard HW filtering, followed by visual manual removal of noise contaminated epochs.
(Goldfine, et al. 2011)	1.Manual rejection of snippets with visible motion artifact 2.Removal of 60Hz line noise 3.Removal of EMG: electromyography and eye movement artifact with ICA
(Groullier, et al. 2007)	Image Artifact Reduction - Adaptive Noise Cancelling to remove repetitive MRI noise. FASTR - takes PCA, sorts by variance they explain, and throws away first few "Since the imaging artifacts are uncorrelated to neuronal activity and of much higher amplitude..." ICA (assumes common mode signals are noise, because spatially different EEG signals are "generated by different (uncorrelated) processes." Filtering using Fourier Transform - multiplied each frequency with the inverse of the power of the noise at that frequency (instead of zeroing which introduces ringing and removal of important info) Average Artifact Subtraction - removing an average of 30 heartbeats, however did not have to deal with ocular and muscular artifacts, nor QRS detection since data was all simulated. Kalman adaptive filtering - puts a motion sensor over the temporal artery and then uses that data to remove ECG from EEG. PCA for ECG removal - the first components of a PCA are assumed to capture most of the variance introduced by the BCG. However, removing too many components deteriorates the EEG data as EEG and BCG are not strictly orthogonal.
(Gwin, et al. 2010)	Visual inspection and removal of noisy channels. ICA was used for noise reduction. Before performing ICA decomposition, time periods of EEG with substantial artifact, based on the z-transformed power across all channels at a given time-point being 0.8, were rejected using EEGLAB. The rejected frames were inspected visually, and regions of 50 accepted frames between any two sets of rejected frames were also rejected. Epochs containing artifacts not related to locomotion (such as eye movements and line noise) were excluded from further analysis using EEGLAB routines that determined the probability of occurrence of each trial by computing the probability distribution of EEG channel signals. Epochs with a probability of occurrence 3 SD from the mean across all epochs were rejected from further analysis. The remaining epochs were averaged to form EEG channel-based ERPs. Additionally, these epochs were multiplied by the ICA unmixing matrix to form IC activity epochs and averaged across epochs to form IC-based ERPs - Heel strike events normalized to identical latencies through time warping of EEG signal according to motion-tracking cameras and pressure sensors.
(Heraz and Frasson 2007)	Removed experiments with insufficient data (not enough participates rated their emotion at that level).
(Ito, et al. 2010)	Time averaging of power spectrum.

(Jung, et al. 1997)	EEG data were recorded at a sampling rate of 312.5 Hz from two midline sites, one central (Cz) and the other midway between parietal and occipital sites (Pz/Oz), using 10-mm gold-plated electrodes referenced to the right earlobe. EEG data were first preprocessed using a simple out-of-bounds test (with a f 5 0 UV threshold) to reject epochs that were grossly contaminated by muscle and/or eye-movement artifacts. Moving-averaged spectral analysis of the EEG data was then accomplished using a 256-point Hanning-window with 50% overlap. Windowed 256-point epochs were extended to 512 points by zero-padding. Median filtering using a moving 5-s window was used to further minimize the presence of artifacts in the EEG records.
(Khare, et al. 2009)	Occipital sensors used to remove EOG
(Khosrowabadi, et al. 2010)	EEG normalized for each subject separately.
(S. Lee, B. Abibullaev and W.-S. Kang, et al. 2010)	Wavelet based de-noising technique also called as the universal threshold estimation method. The objective of wavelet de-noising algorithm is to suppress the noise part of the signal. Universal threshold estimation method uses a fixed for threshold. THRuni = SIGMA (sqrt(2log(N))) Where N is the length in samples of time-domain signal, SIGMA is standard deviation of noise, and log is a natural logarithm. Using these methods, calculated the wavelet coefficient values of each wavelet transform functions.
(Ma, et al. 2012)	Primarily Spatial ICA, however a muscle artifact identifying "Spectral Ratio" function is proposed equaling log(HF/LF) where HF is the average spectral power from 40 to 60 Hz, and LF is the minimum spectral power of the smoothed (averaged over time) frequencies from 4 to 30 Hz. This presumes that muscle signals do not have the same Low Pass filtering that EEG signals have passing through layers of brain matter prior to reaching the electrode.
(MacLean, Arnell and Cote 2012)	Visually remove EEG epochs with artifacts.
(Mostow, Chang and Nelson 2011)	Averaged data over the time interval of each stimulus.
(de Munck, et al. 2013)	Subtract a template of the noise artifact (aligning EEG's exactly and averaging them to get the high quality template requires EEG at 2K-5K Hz sampling) but only good for mechanical noise from fMRI. Heart beat noise can also use template subtraction (median instead of average to avoid extreme outliers), for template subtraction. ICA can be used, but it assumes artifacts are statistically independent from EEG, and require visual inspection to select components. Hierarchical clustering of precise time shifts is used, and small clusters are deemed to be noise and are removed. In this way, it is possible to get groups of time shifted noise to effectively subtract
(Murugappan 2011)	6th order Butterworth used to filter out noise.
(Nagashino, et al. 2002)	No noise removal.
(Palaniappan 2006)	Order 5 elliptical filters were used - one forwards and once backwards to account for phase - at each of the bands. Spatial symmetry at each of the bands is the metric used in this experiment.
(Safieddine, et al. 2012)	ICA and canonical correlation analysis across all electrodes, and empirical mode decomposition (EMD) and wavelet transform on each individual electrode are compared. Here, "noise includes EMG and EKG, but also EEG not related to interictal spikes from epilepsy." EMD performed best with extremely muscular contaminated noisy data, however this varied greatly depending on whether the data was very contaminated with noise, and whether the signal to be cleaned was originating from a single electrode position, or multiple ones.

Table 16 Survey of various EEG noise decontamination methods used by researchers

Ref	Windowing Parameters
(Wilson and Bracewell 2002)	5 second windows (rectangular envelope presumably) with no overlap described.
(Khare, et al. 2009)	10 second windows (rectangular envelope presumably)
(Jung, et al. 1997)	1.2 second windows, using a Hanning-window and then zero padded to 2.4 seconds. Each windows had a 50% overlap with the prior window.
(Palaniappan 2006)	0.5 second windows (rectangular envelope presumably) with no overlap described.
(Ito, et al. 2010)	60 second windows (rectangular envelope presumably) with no overlap described.
(Khosrowabadi, et al. 2010)	2 second windows (rectangular envelope presumably) with no overlap described.
(Murugappan 2011)	Entirely variable length windows (rectangular envelope presumably) with no overlap described.

Table 17 Survey of EEG data windowing parameters used by researchers

Research Paper	Delta	Theta	Alpha	Beta	Gamma	Band above Gamma
(Sandberg, et al. 2010)		3.9-12.7 Hz Theta and Alpha				
(Wilson and Bracewell 2002)	0.5-2 Hz	3-7 Hz	8-12 Hz	12-14 Hz “spindles”		
(Khare, et al. 2009)			8-13 Hz			
(Chakraborty, et al. 2009)	0-4 Hz	4-8 Hz	8-10 Hz lower 10-12 Hz upper	12-16 Hz Beta 1 16-24 Hz Beta 2	24-32 Hz	
(De Vico Fallani and al 2008)		3-6 Hz	7-12 Hz	13-29 Hz	30-40 Hz	
(Subasi, et al. 2005)	1-4 Hz	4-8 Hz	8-13 Z	13-30 Hz	> 30 Hz	
(Palaniappan 2006)	0-3 Hz	4-7 Hz	8-13 Hz	14-20 Hz	24-37 Hz	
(Murugappan 2011)	0-4 Hz	4-8 Hz	8-16 Hz	16-32 Hz	32-64 Hz	Noise
(Mostow, Chang and Nelson 2011)	1-3 Hz	4-7 Hz	8-11 Hz	12-29 Hz	30-100 Hz	Gamma+
(Berka, et al. 2004)			8-12 Hz			
(Pasqualotto, Federici and Belardinelli 2012)			8-14 Hz	18-26 Hz		
(NeuroSky Co. Feb 11, 2012)	0.1-3 Hz	4-7 Hz	8-12 Hz	12-15 Hz Low 16-20 Hz Mid 21-30 Hz High		
(Uusberg, et al. 2013)			8-13 Hz (upper Alpha 11-15 Hz)			

Table 18 Survey of different EEG band definitions used by researchers

Citation	EEG Application	Participants
(Bakera, et al. 2010)	Alertness (simulation defending against incoming enemy planes), cognitive task (response rate in identifying the number "5" out of digits presented), and memory (recognizing memorized images) is detected using EEG data and "B-Alert" pattern recognition software to detect alertness.	45 image learning task 13 for Warship commander task
(Belle, Hobson Hargraves and Najarian 2012)	Engagement and attention detection towards lengthy videos.	12
(Chakraborty, et al. 2009)	When presented with an audio-visual stimulus designed to evoke an emotion, EEG data (from position F3 or F4) is used to predict the user's facial expression. 50 volunteers were shown 60 audio-visual clips, covering a range of six different emotions (anxiety, disgust (anger), fear, happiness, sadness, and relaxation)	50 used to classify clips. May also be the same number used for EEG. Later additive Gaussian noise is used to create a data set of "100 subjects"
(Crowley, et al. 2010)	Volunteers wearing the headset are asked to do an increasingly stressful (faster) Stroop Test (name the color of the word, even though font color is green and word says "red" etc.) - every time the speed was increased, EEG registered dips in "meditation" (lower relaxedness). Similarly, the Tower of Hanoi problem was given to volunteers three times in a row - it seems tricky but one you learn the stacking algorithm it's easy to solve. EEG data was able to show trends in lowered stress	20 for one test, and 17 for another.
(De Vico Fallani and al 2008)	A measure of "network efficiency" based on EEG data (with manually removed ocular artifacts) is used to predict whether a short video clip will be remembered or not several days after viewing. Study shows when a video clip is forgotten, there is more network efficiency (global efficiency as measured by efficiency index), and less efficiency when the clip is remembered, especially with regard to the Apha, Beta, and Gamma bands.	19
(Gargiulo, et al. 2010)	Testing of a new dry-electrode EEG data collection system	3 and 8
(Goldfine, et al. 2011)	Can EEG Frequency analysis show that the brain injured who are otherwise unable to communicate still have frequency variations (can demonstrate awareness) when asked to mentally imagine motor and spatial navigation tasks	3 brain injured and five health control subjects
(Green and Najarian 2007)	Explore a person's emotional regulation with their EEG response towards training situations.	19
(Groullier, et al. 2007)	Finding an optimal EEG artifact removal technique finds that the optimal artifact removal technique should be chosen according to whether one is interested in fast (N10 Hz) vs. slow (b10 Hz) oscillations or in evoked vs. ongoing activity.	1 participant, and then simulated EEG signals created from that.
(Gwin, et al. 2010)	Taking EEG while walking or jogging on a treadmill and visually identifying "oddball" stimulus.	8 participants (actually 6 because 2 got bad data when running)
(Heraz and Frasson 2007)	Predicting the Three Major Dimensions of the Learner’s Emotions from Brainwaves	17

(Ito, et al. 2010)	Simple EEG Headset (earlobe referential and Fp1) is used to collect data while participant listens to music segments (which they later rate as "Matches Mood", "Does not match mood", and "Borderline"). Separately, the psychological score of the participant is known from a 50 question exam (rating the person in the 5 areas of "Critical parent", "Nurturing parent", "Adult", "Free Child", and "Adapted Child".) The authors then compare the false-detection accuracy of the EEG classification system of "mood matching" to the personality trait of the participant, stating that false-detection may be related to the person's personality.	7
(Jung, et al. 1997)	EEG data from two scalp sites ("two midline sites, one central (Cz) and the other midway between parietal and occipital sites (Pz/Oz), using 10-mm gold-plated electrodes referenced to the right earlobe.") are used to predict a volunteer's alertness (as measured by lapses in auditory and visual sonar detection by trained Navy volunteers.)	15
(Khare, et al. 2009)	Try to use EEG data to classify mental state. Five mental states used were all wheny lying still: 1- relaxed, 2 - imagine moving your right hand, 3- watching a figure being rotated (imagining it as well), 4 - trivial multiplication (2x3), and 5 - non-trivial multiplication (49x78).	10
(Khosrowabadi, et al. 2010)	Participants presented with pictures and music to elicit four basic emotions of calm, happy, sad and fear, and they are also questioned using a self-assessment manikin as to the emotion they felt. The emotion is mapped onto a Valence-Arousal emotion plane. EEG is then measured at points F3, F4, T7, C3, CZ, C4, T8, P3, and P4 - to see if emotion can be pattern-detected.	31 from which 26 were manually selected.
(S. Lee, B. Abibullaev and W.-S. Kang, et al. 2010)	Children between the age of 10 and 13 years with ADHD and parental consent were assessed for mental retardation using KEDI-Wechsler Intelligence Scale for Children, and then EEG data was collected at 25.6 Hz sampling from 2 frontal lobe points. Participants were asked to relax, count, and perform mental tasks. The EEG data was used to pattern-recognize if the child had only ADHD, or also had mental retardation.	13
(Lisetti and Nasoz 2004)	EEG was not used. Other physical measures "galvanic skin response [GSR], heart rate, and temperature" were used to predict emotional state of sadness, anger, amusement, fear, and surprise through watching clips from popular movies. Also collected were repeated values of the users responses to relaxing music before the presentation of each clip.	14 (minus "data loss for the data of some participants during segments of the experiment")
(Ma, et al. 2012)	Noise reduction in multi-channel EEG data - an automated method.	12 and 16 subjects with 48 5-min data collections per subject over several days / activities.
(MacLean, Arnell and Cote 2012)	Some people have a bigger Beta versus Alpha, even when they are not engaged in a goal-directed task. These people have a shorter "attentional blink" which is if they are serially presented with targets, noticing T1 may cause them to miss noticing T2 if it occurs less than half a second after T1 (that time being the AB).	Two experiments; 27 = 30 minus 3 excluded (EEG error or bad perf) and then 29 = 38-9
(Mostow,	Easy and hard reading for meaning tasks assigned and EEG data compared for adults and	15 (6 adult, 9 child) after

Chang and Nelson 2011)	children.	excluding 6 for poor EEG data quality
(de Munck, et al. 2013)	Is it possible to subtract fMRI electromechanical noise, as well as noise caused by the heartbeats that move the EEG wires ever so slightly by time shifting and subtracting the noise data?	29 healthy in one scanner and 15 with epilepsy in a different scanner
(Murugappan 2011)	5 Videos elicit emotions (disgust, happy, fear, surprise and neutral) as selected by a college student panel from 115 "international standard emotional clips". Separate volunteers wear 64 electrode EEG hats and the data is used to detect which emotion. NOTE: the author states "One of the limitations in this area of research is lack of international standard data base. Hence, all the researchers are reporting their results according to their own set of data bases. The accuracy of classification depends on the number of electrodes used, method of emotional stimuli used..."	20
(Nagashino, et al. 2002)	EEG used to discriminate whether a volunteer in a relaxed state has their eyes open or closed.	One healthy female
(Palaniappan 2006)	EEG data is used to determine the mental activity of the volunteer from a series of mental tasks including Geometric figure rotation, multiplication, writing a letter to a friend, visual counting, and resting.	4
(Rebolledo-Mendez, et al. 2009)	Use of EEG Attentiveness measuring device to measure whether people are attentive during an assessment exercise. In this case, a single dry electrode with ear ground is used and volunteers interact with a Second Life avatar that asks multiple choice questions. EEG attention rating from 0 to 100 is correlated with speed/accuracy of response and user self-assessed attentiveness levels. There was no training going on, questions were targeted at Computer Schience, and volunteers all were majors in CS.	40
(Safieddine, et al. 2012)	Is it possible to remove muscle contamination from EEG signals to better find epileptic Interictal Spikes	none-artificial data used. System was then tested on one participant with epilepsy
(Uusberg, et al. 2013)	Instead of just viewing the Alpha band as passive idling indicator (or a band whose power is increased when the participant is resting or inattentive) this research examines it's role as an inhibitor, this research tries to show that the alpha band is present in varying amounts depending on the affective stimulus (pictures from the International Affective Picture System). It was found that the prevalence of details in the picture affected Alpha, and as such cannot be discounted when comparing affective stimulation results. Also it was found that aversive (powerfully unpleasant) images generated high Alpha power, compared to rest and other images.	95 reduced down to 73 based on EEG data and left-handedness

Table 19 Survey of number of participants used in EEG research studies

Citation	Validation Method	Validation Style
(Bakera, et al. 2010)	Statistical - Analysis of Variance (ANOVA) by Group	Statistical
(Belle, Hobson Hargraves and Najarian 2012)	Leave One Out (LOO)	Participant Cross Validation with 12 participants
(Chakraborty, et al. 2009)	Data across all volunteers (60, 20, 20% train, validate, test) until got a minimum mean squared error.	Cross Validation including data from all participants
(Crowley, et al. 2010)	Human observer compared to EEG measured stress levels	Manual
(De Vico Fallani and al 2008)	"ANOVA – with a significance level of 0.05 was employed. In particular, the main factors of the two ANOVAs performed for the global- and local-efficiency indexes were the factor TASK with two levels (spot-forgotten, FRG; spotremembered, RMB) and the factor BAND with four levels related to the frequency bands adopted (theta, alpha, beta and gamma). Post hoc comparisons between different levels of the investigated factors were performed with the Scheffe's test, at the 0.05 of statistical significance."	Statistical
(Gargiulo, et al. 2010)	Comparing covariance of simultaneous wet and dry electrode data	Statistical
(Goldfine, et al. 2011)	"Analysis Across All Frequencies 1.Fisher's linear discriminant (FLD) created for each channel from log power spectra of task and rest snippets using 2Hz-wide bins between 4 and 24 Hz 2.FLD p-value calculated from 1000 shuffles of task and rest 3.FDR applied to FLD p-value to correct for tests of all channels"	Statistical
(Green and Najarian 2007)	The test of statistical significance of the features was performed using ANOVA to determine the statistical significance of each of the extracted feature across post-study analysis groups.	Statistical
(Groullier, et al. 2007)	Comparing the simulated EEG data with the noise added then noise removed signals	Statistical
(Gwin, et al. 2010)	Visual inspection of performance of noise removal algorithms	Manual
(Heraz and Frasson 2007)	"ANOVA showed correlation of four brainwave parameters (power of delta 0-4, theta 4-8, alpha 8-12 beta 12+) and three emotional dimensions. 10-fold cross validation was used to test machine learning algorithms (using data from all volunteers)."	Statistical and Cross Validation including data from all participants
(Ito, et al. 2010)	LOO using all the EEG feature vectors	Participant Cross Validation with 7 participants
(Jung, et al. 1997)	2-fold cross validation on features from the same subjects were used.	Cross Validation including data from all participants
(Khare, et al. 2009)	60% / 40% split of data used for training / testing from a mixture of data from all participants.	Cross Validation including data from all participants
(Khosrowabadi, et al. 2010)	5-fold cross validation. Not clear if this was separated by participant or not…	Cross Validation including data from all participants
(S. Lee, B. Abibullaev and W.-S. Kang, et al. 2010)	Simply compared the clustering results between 4 different wavelet mother shapes. No cross validation, just comparison against expert analysis of subjects.	Manual

(Lisetti and Nasoz 2004)	When using kNN, training and test sets came from mixed pool of subjects. When using discriminant analysis, all samples used. For back-prop, all samples were used for both training and test.	Cross Validation including data from all participants
(Ma, et al. 2012)	Clustering is used to create a set of 10 "major categories of muscle components" and these synthesized muscle components are used for simulating EEG+muscle signals. The algorithms were then used on real EEG data, and pre and post filtered data was visually compared. Future work still remains if this is valid to retain important EEG information.	Manual
(MacLean, Arnell and Cote 2012)	3 (band) x 4 (region) ANOVA against the AB values. This was used to show the high correlation of Beta at rest with low AB values.	Statistical
(Mostow, Chang and Nelson 2011)	Cross validated per-participant (leaving out one epoch) and across participants (LOO). Accuracies were averaged.	Participant Cross Validation with 15 participants
(Murugappan 2011)	Majority vote kNN using k=2 to 8 (k=6 was the winner). Five-fold cross validation used however mixtures include all subjects.	Cross Validation including data from all participants
(Nagashino, et al. 2002)	Neural network learns EEG time sequence, and the RMS error between NN predicted EEG time sequence, and actual is measured.	Statistical
(Palaniappan 2006)	Ten-fold cross validation was used taking from the data from all subjects.	Cross Validation including data from all participants
(Rebolledo-Mendez, et al. 2009)	Pearson correlation between self-reported attention levels against device measured attention level.	Statistical
(Safieddine, et al. 2012)	Normalized Mean Squared Error in ability to extract original EEG signal.	Statistical
(Uusberg, et al. 2013)	ANOVA and ANCOVA (analysis of covariance)	Statistical

Table 20 Survey of validation methods used in EEG research studies

Appendix B Consent Form

RESEARCH SUBJECT INFORMATION AND CONSENT FORM

TITLE: Physiological Factors in a Challenging Task

VCU IRB NO.: HM13618

This consent form may contain words that you may not be completely clear to you. Please ask the study staff to explain any words that you do not clearly understand. You may take home an unsigned copy of this consent form to think about or discuss with family or friends before making your decision.

PURPOSE OF THE STUDY

This study seeks to understand the ways and level of your engagement in a challenging cognitive performance activity relates to a variety of physiological factors. The research project involves a laboratory session taking about 2 hours to complete. Participation is voluntary, and all responses will remain strictly confidential.

DESCRIPTION OF THE STUDY AND YOUR INVOLVEMENT

If you decide to be in this research study, you will first be asked to sign this consent form after you have had all your questions answered and understand what will happen in the study. After you sign this consent form you will be asked to perform a cognitive task and complete a small number of additional questionnaires. Research staff will be present during completion of the cognitive task. Physiological measures of electrocardiogram (ECG), galvanic skin response (GSR), Heat Flux, Temperature, and electroencephalogram (EEG) may be collected during the task through the use of electronic sensors on the skin. The physiological measures may be obtained through an armband device (smaller than a blood pressure cuff) and a small cap to be worn on your head. The armband will measure your ECG (heart rate and signals), GSR (electrical conductance of the skin), temperature and heat flux. The cap will measure you EEG or brain waves. You will be video recorded during the task. As soon as certain anonymous information is coded from it, this video record will be destroyed.

The sampled physiological data will be analyzed for the purposes of this research only. After the completion of this study the data samples will be destroyed. There will not be a way to link these samples to your name; only an anonymous ID number will be generated.

You do not have to answer any questions, or participate in any activities, you do not wish to. You may withdraw from the study at any time, without penalty.

Significant new findings developed during the course of the research which may relate to your willingness to continue participation will be provided to you.

RISKS AND DISCOMFORTS
The physical risks involved in this study are minimal. The procedure is completely noninvasive.

The risks are not greater than the risks associated with daily living. However, if participating in this study causes you to feel upset or you become concerned about your psychological state or your current life situation, the study staff will provide you with contact information for resources available on the VCU campus that can help you address these issues, including:

Center for Psychological Services and Development, which offers counseling services on a sliding fee scale; phone 828-8069.

VCU Department of Psychiatry, which offers acute care services; phone 828-2000.

Fees for treatment will be billed to you or to appropriate third party insurance.

BENEFITS TO YOU AND OTHERS
You may not benefit from participation in this study.. However, the study is likely to yield generalizable knowledge to further society's understanding of the cognitive processes under investigation.

COSTS
There are no costs for participating in this study other than the time you will spend completing the task and filling out questionnaires.

ALTERNATIVES
The alternative to participating in this study is to not participate.

CONFIDENTIALITY
Potentially identifiable information about you will consist of a video record. That is, your study session will be video recorded, but your name will not be recorded. The video record and all other data are being collected only for research purposes. Your data will be identified only by an anonymous ID number that is generated, not by your name or other personally identifying information. The video record and the notes on them will be stored in a data collection computer and in an access controlled secure server that is present within a locked laboratory at VCU. After the information from the video is coded, it will be destroyed. Access to all data and samples will be limited to study personnel. A data and safety monitoring plan is established.

We will not tell anyone the answers you give us; however, information from the study and the consent form signed by you may be looked at or copied for research or legal purposes by Virginia Commonwealth University. What we find from this study may be presented at meetings or published in papers, but neither your name will not ever be used in these presentations or papers. Your participation in this study will not impact your status at your University, nor will your involvement in this study be disclosed to any coworkers or supervisor(s) you may have.

GENETIC TESTING
This research does not involve genetic testing

IF AN INJURY HAPPENS
Virginia Commonwealth University and the VCU Health System (also known as MCV Hospital) do not have a plan to give long-term care or money if you are injured because you are in the study. If you are injured because of being in this study, tell the study staff right away. The study staff will arrange for short-term emergency care or referral if it is needed. Bills for treatment may be sent to you or your insurance. Your insurance may or may not pay for taking care of injuries that happen because of being in this study.

VOLUNTARY PARTICIPATION AND WITHDRAWAL
You do not have to participate in this study. If you choose to participate, you may stop at any time without any penalty. You may also choose not to answer particular questions that are asked in the study. Withdrawal from the study will not affect you present or future University relationship.

Your participation in this study may be stopped at any time by the study staff without your consent. The reasons might include:

the study staff thinks it necessary for your health or safety;
you have not followed study instructions;
administrative reasons require your withdrawal.

QUESTIONS
In the future, you may have questions about your participation in this study. If you have any questions, complaints, or concerns about the research, contact:

Dr Rosalyn Hobson
Virginia Commonwealth University
School of Engineering
601 West Main Street, Suite 331
PO Box 843068
Richmond, VA 23284-3068
Phone: (804) 828-8308
Fax: (804) 827-0006

If you have any questions about your rights as a participant in this study, you may contact:

Office for Research
Virginia Commonwealth University
800 East Leigh Street, Suite 113
P.O. Box 980568
Richmond, VA 23298
Telephone: 804-827-2157

You may also contact this number for general questions, concerns or complaints about the research. Please call this number if you cannot reach the research team or wish to talk to someone else. Additional information about participation in research studies can be found at http://www.research.vcu.edu/irb/volunteers.htm.

CONSENT

I have been given the chance to read this consent form. I understand the information about this study. Questions that I wanted to ask about the study have been answered. My signature says that I am willing to participate in this study. I will receive a copy of the consent form once I have agreed to participate.

________________________ ________________________ __________

Participant name printed ***Participant signature*** ***Date***

____

Name of Person Conducting Informed Consent/Witness (Printed)

____ ________________

Signature of Person Conducting Informed Consent/Witness **Date**

____ ________________

Investigator Signature (if different from above) **Date**

Appendix C Triangulation with Qualitative Data

As described in the background chapter, for the purposes of this research, the affective state of the participant watching the short training video is deemed to be attentive when the participant has a high positive affect, including feelings of satisfaction, engagement, interest, and involvement. The current research used a questionnaire after the video to let the participant document these thoughts and feelings in their own words. In addition to taking quantitative measurements of EEG signals and timing of events in the videos versus timing of attentiveness measures, it is also useful to examine qualitative data relating to attentiveness to check if it matches what is in the quantitative data. The post video questionnaire asked participants to put in their own words what they were thinking, and how they felt during different parts of the video. These are called the PostQA. In addition, the investigator was sitting in the room and made observations if anything out of the ordinary took place alongside the timing data. These are called the timing notes. It is possible to compare these to the attentiveness measure of the individual participant and see if the quantitative data corresponds to the respective qualitative data.

In order to perform this triangulation with qualitative data, a participants timing notes and PostQA information is shown just below a graph of their attentiveness measure. The attentiveness measure in the diagrams below are from the DFT system, without the use of a baseline (so as to highlight any participants who had attentiveness measures that varied greatly from the other participants). The data was smoothed with a 40 second window and was measured over 400 seconds of half boring and half attentiveness inducing training video segments – minus 20 seconds at the front and back end for the smoothing window. A cross section of largely randomly selected participants are shown in the figures and tables below, although some participants were purposely selected because they had unique situations during the experiment, or because their EEG had been manually removed from the participant pool due to excessive EEG noise. This was done on purpose to see if the triangulation provided meaningful results.

In each section below, an individual participant is described with a table of information that includes:

- Participant - The number of the participant + 100 (to distinguish from the pilot study participants that started with number 1)
- Self-ID? - A 0 if the participant did not self-identify in the PreQA, a 1 otherwise.
- New PreQA? - A 0 if the participant used the old PreQA questionnaire, and a 1 otherwise.
- Video Used - a number indicating the video viewed, 1, 2, or 3 signifying Video1, Video2a, or Video2b respectively
- Noisy EEG? - a subjective qualitative measure of whether or not the EEG data visually looked so noisy as it was removed from the training data sets (manual noise removal algorithm)
- Clean EEG? - a subjective qualitative measure of whether or not the EEG data visually looked particularly clean with regard to showing a lack of noise spike events.

After the participant description table, an attentiveness graph over time is shown. This graph is produced by output of the DFT system created in this research. The graph has data from the boring segment of the video on the left (time markers are 0 to 200 seconds) and data from the attentiveness inducing segment of the video to the right (time markers are 200 to 400 seconds).

Beneath the graph is a second table which shows, in the participants own words, how they answered the questions "How did you feel?" and "What were you thinking?" for the beginning, middle, and end of the boring video segment, and for the beginning, middle, and end of the training video segment. Values under 0.5 show boredom or inattentive EEG classification, whereas values over 0.5 show attentiveness indicators. Without the use of baselines, one would expect that the total data may be shifted up or down for that particular participant, but baselines are not used for the reasons explained above. Zero padding for seconds 0-20 and seconds 380-400 are because of the 40 second smoothing window used.

Finally, in each section below, any relevant information from the investigator's timing notes is described, as well as summary analysis of the triangulation of qualitative data for that participant.

Attentiveness Graphs of some Typical Participants

These are presented in no particular order

Participant Number 3

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
103	0	0	1	0	0

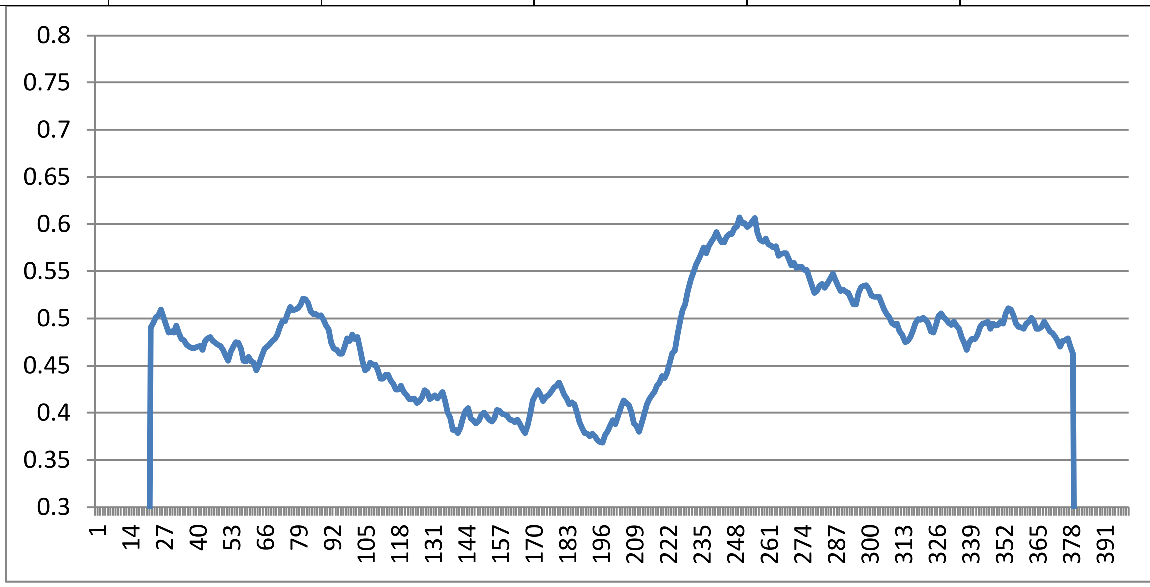


Figure 51 Attentiveness Graph (no baselines used) for Participant 3

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Oh, OK. It was pretty basic.	I got bored. I wondered how much longer.	Ready to move on.	I was making sure I was prepared	I was trying t pay attention that I did it the right way.	I felt relieved.
Thinking		Drift off. Look around at computer and other stuff	Wondered how much longer	Make sure I was ready to listen; preparing myself; alert.	Not really anything. Just making sure I folded paper exactly as the video	

Table 21 Thoughts and Feelings as Reported By Participant 3

There were no EEG connectivity issues for this participant who viewed Video1 and was observed to be on task for the training portion (folding paper, etc.) The student sat diagonally so they could see the investigator in the reflection of the laptop screen (investigator sat slightly behind and to the right of the participant). There was one point during the boring video when the investigator’s cellphone could be heard to buzz loudly (it was resting on a hard table instead of in the investigator’s pocket). This took place just before the 80 second mark, and triangulates well with an attentiveness spike centered at that point (averaging attentiveness using values 20 seconds before and after cause this to appear spread out form 60 seconds to 100 seconds). Otherwise the shape of the attentiveness graph looks very similar to the ensemble average of participants who watched Video1.

Participant Number 11

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
111	0	0	1	0	1

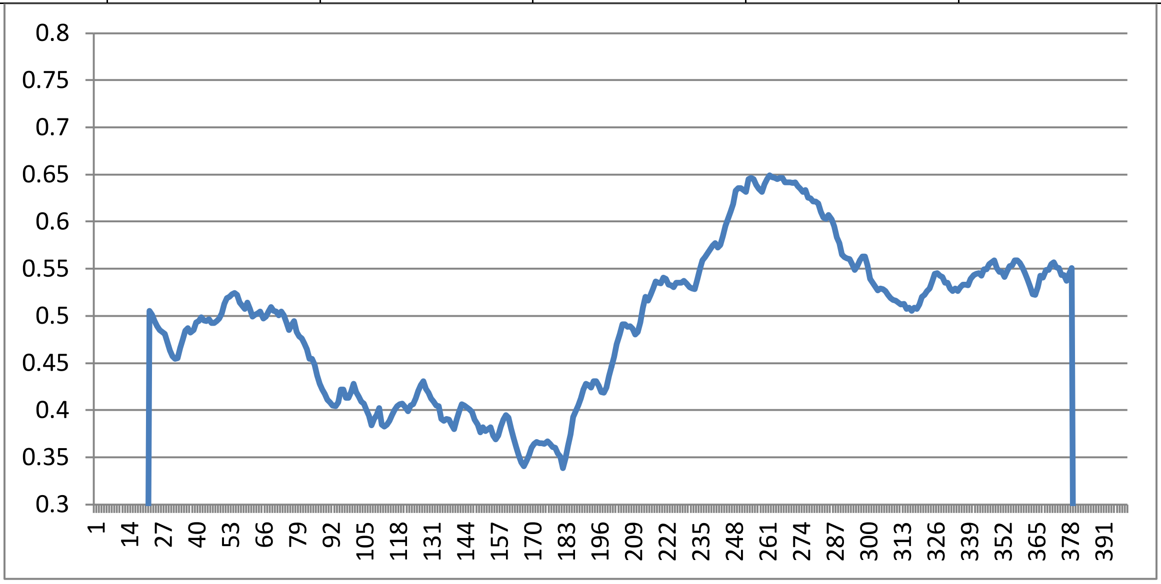


Figure 52 Attentiveness Graph (no baselines used) for Participant 11

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	I thought it was going to be only for a second	Wondering if it was broken. I was watching the time	Really bored. Wanted it to be over.	Anxious, what will it be?	Like I was following instructions. I rather follow, rather watch and do.	I thought my writing was pretty.
Thinking	I was excited, what will it be about, helping	I hope I'm doing it right.	Bored	Glad the candle video was over	I thought it was better. I like markers.	Thinking I was going to mail it to someone.

Table 22 Thoughts and Feelings as Reported By Participant 11

There were no EEG connectivity issues and the participant watched Video1 and was observed to be on task during the training segment (folding paper, etc.). There was one technical issue in that the video had the time progress bar at the bottom moving from left to right (normally the progress bar disappears shortly after the video starts, but the investigator had mistakenly positioned the mouse causing this problem). The shape of the attentiveness graph is typical and similar to the ensemble average of participants who watched Video1. There is a bit of attentiveness at the start of the boring half, which wanes and triangulates well with the participants self-described increasing boredom. The attentiveness towards the training video seems to start ramping up early. There is a point at the end of the boring half, but a few seconds before the training half starts, where the words "relax and focus on your breathing" disappear, but the candle still is shown. The ensemble average shows attentiveness increasing at this point, but this particular participant seems to have attentiveness start the ramp up earlier. This seems to triangulate with the fact that they can see the progress bar nearing the middle of the screen, and knowing beforehand that there will be a boring video followed by a short training video.

Participant Number 48

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
148	1	1	3	0	0

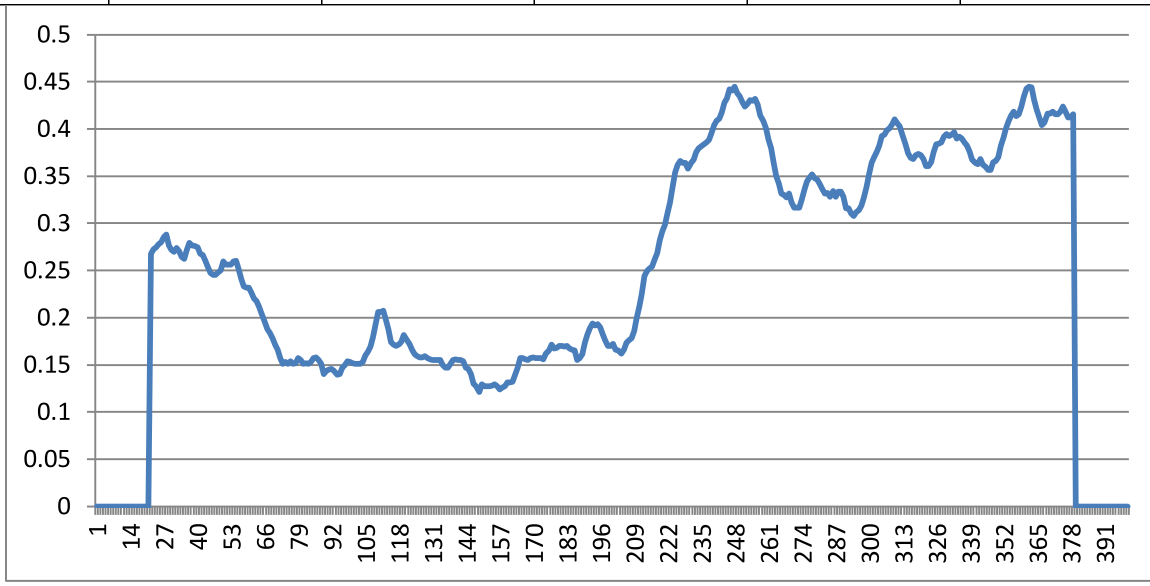


Figure 53 Attentiveness Graph (no baselines used) for Participant 48

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Relaxed	Relaxed	Focused	Calm	Relaxed	Focused
Thinking	Food	School	Nothing	2	4	5

Table 23 Thoughts and Feelings as Reported By Participant 48

There were no EEG connectivity issues, and the participant watched Video2b with the candle first, and the remembering the three numbers for the second half. The participant was observed to be largely sitting still throughout (moved feet at one point) and remembered all three numbers after the video was over. Aside for the need of the baselines to shift the data upwards, this looks like a typical EEG for this research.

Participant Number 25

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
125	0	1	1	0	1

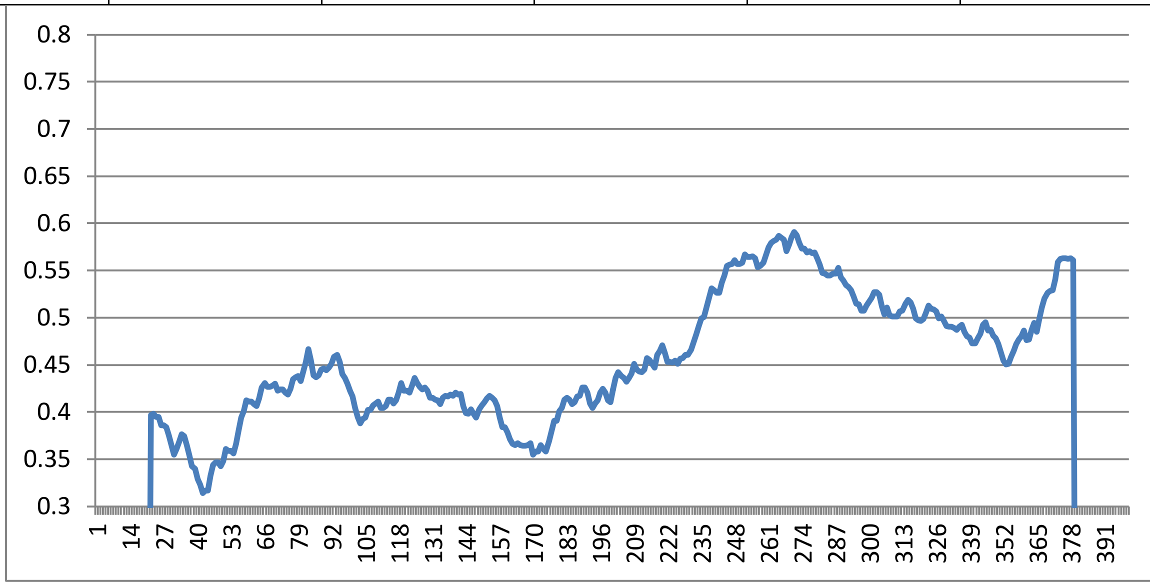


Figure 54 Attentiveness Graph (no baselines used) for Participant 25

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Curious about what I was going to do.	Lots of unrelated thoughts, candle reminded me of the wii sports game	Wondering what the outcome of the research would be.	Uncertain but curious	Though about how easy the research is	At ease but still curious of outcome.
Thinking	Wii game, balance	Not very rested, felt sleepy	Trying not to let my eyes close	Unsure of what paper and supplies were for.	Felt like kindergarten	It was easy.

Table 24 Thoughts and Feelings as Reported By Participant 25

No EEG connectivity issues for this participant who watched Video1 and was observed to be on task (folding paper, etc.) during the training half of the video. The shape of the attentiveness graph is similar to the ensemble average of participants who viewed Video1 with some subtle differences. There seems to be a rise in attentiveness between the middle and the end of the boring half of the video. This triangulates with the comment the participant made with the

PostQA where they said they were thinking “lots of unrelated thoughts.” Another difference is the attentiveness waning at the end of the training video segment, but then spiking up again. The waning triangulates with the comments regarding the training being “easy” and “like kindergarten.” The upwards attentiveness may be related to the comment “curious about the outcome” which the participant used to describe their own feelings at the end of the training video.

Participant Number 2

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
102	0	0	1	0	1

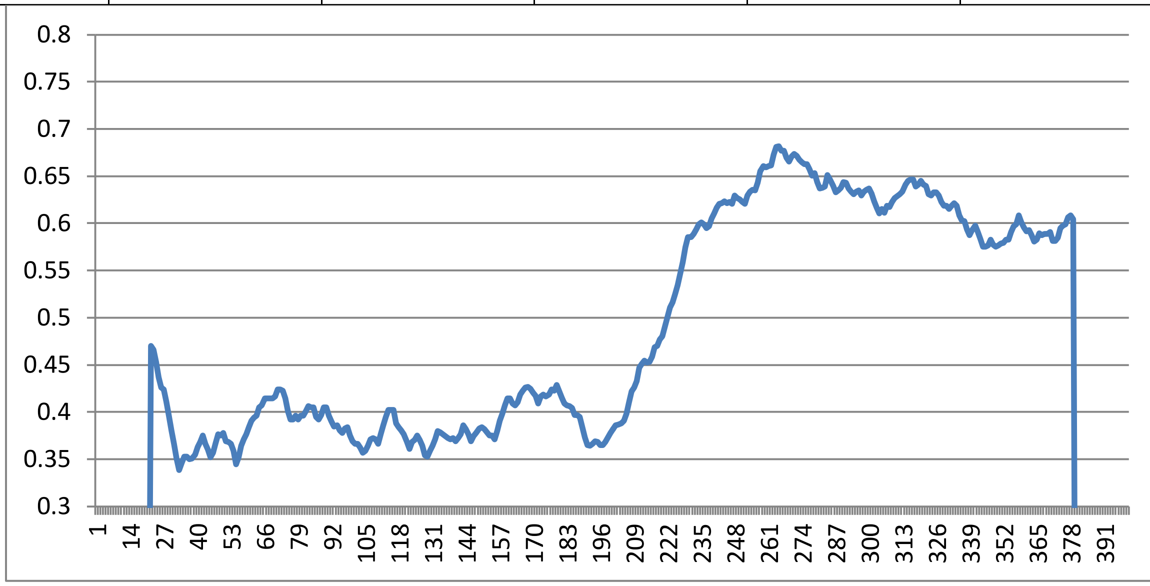


Figure 55 Attentiveness Graph (no baselines used) for Participant 2

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Kind of shut my eyes for a few seconds	Just watched it	I don't know how to explain. My heart was racing fidgety	I found it hilarious	Calm	Yeah!
Thinking	Asking teachers questions and random things	Thinking if device can read my mind.	Said the word "echo" in my mind, having fun with it	I did this wrong.	I did this wrong. Oh well, I messed up.	I was thinking about drawing & coloring. I wanted to draw

Table 25 Thoughts and Feelings as Reported By Participant 2

There were EEG connectivity issues with this participant who had to sit closer to the table to narrow the distance to the BlueTooth wireless receiver. The participant seems to indicate that they were attentive during the boring first half of the video with statements describing various thoughts going through their mind, but this registers only as variability in attentiveness, not a significant major trend that can be triangulated with any accuracy. The entire attentiveness graph seems to be similar to the ensemble average of participants who watched Video1

Participant Number 26

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
126	0	1	1	0	1

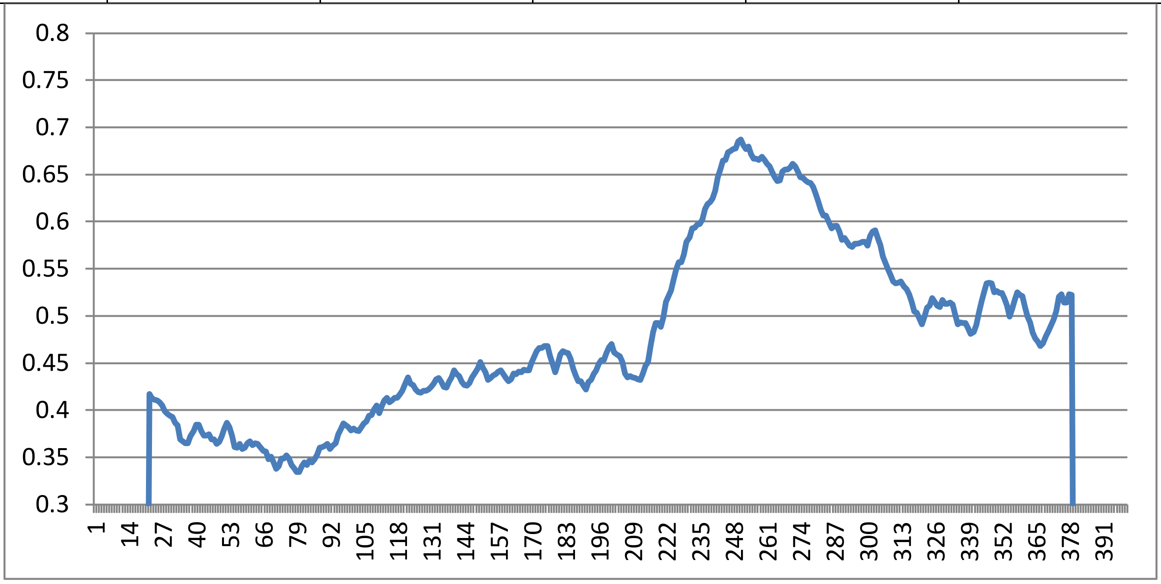


Figure 56 Attentiveness Graph (no baselines used) for Participant 26

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Calm	Relaxed	Relaxed and a little bored	Relaxed	Curious	I expected more.
Thinking	This is all of the video?	Good thing I am not sleepy.	When will this end?	Following directions	This is too easy	I kept getting ahead of the directions

Table 26 Thoughts and Feelings as Reported By Participant 26

No EEG connectivity issues for this participant who watched Video1 and was observed to be on task (folding paper, etc.) during training video. Attentiveness indicator seems fairly similar to ensemble average of other Video1 watching participants. The investigator notes that the participant laughed towards the first part of the training video, when the paper is first folded. Attentiveness seems to wane quickly after that, which triangulates with the comments from the participant saying that they “expected more” and that it was “too easy.”

Participant Number 36

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
136	1	1	1	0	0

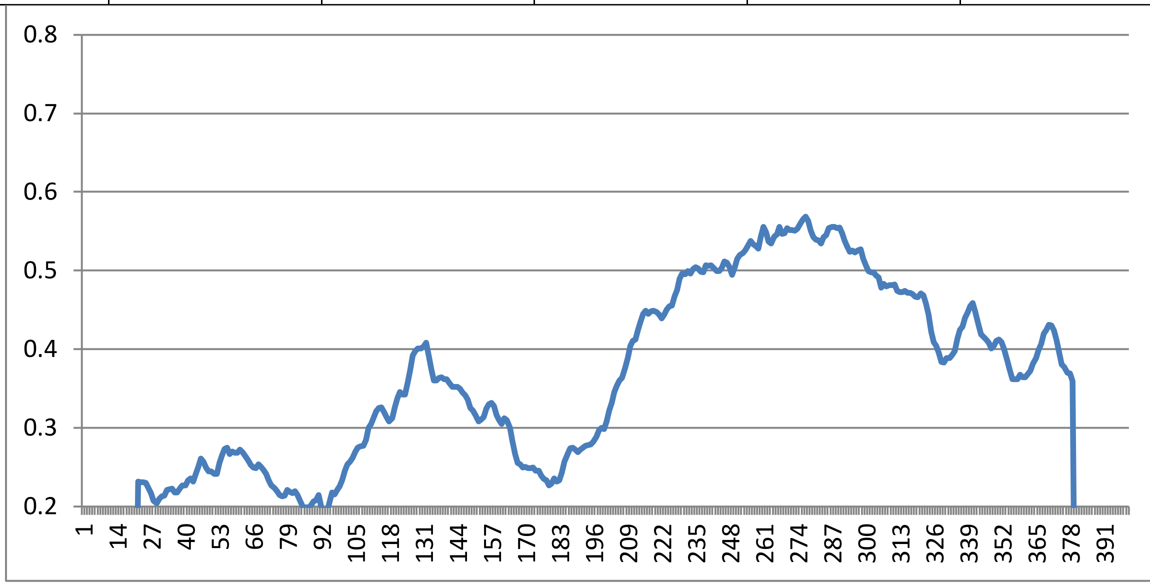


Figure 57 Attentiveness Graph (no baselines used) for Participant 36

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Curious	Relaxed	Bored	Curious	Interested	Accomplished
Thinking	What kind of video is this	Is there anything to the video besides breathing?	How long is this video?	What will the purpose be?	It was a little slow paced	How simple it was

Table 27 Thoughts and Feelings as Reported By Participant 36

There were no EEG connectivity issues as this participant watched Video1 and was observed to be on task during the training (folding paper, etc.). The investigator documented that the participant smelled strongly of alcohol on their breath, but no formal blood alcohol test was performed, nor were other symptoms observed, so this is purely a subjective notation. The attentiveness graph is fairly characteristic of the ensemble average of all Video1 observing participants, except for the very low attentiveness even from the very beginning of the boring video, which is uncommon. On the training half (the right side of the graph) attentiveness peaked and then waned showing triangulation with the participants comments about “how simple it was” and “slow paced.”

Participant Number 41

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
141	0	1	1	0	1

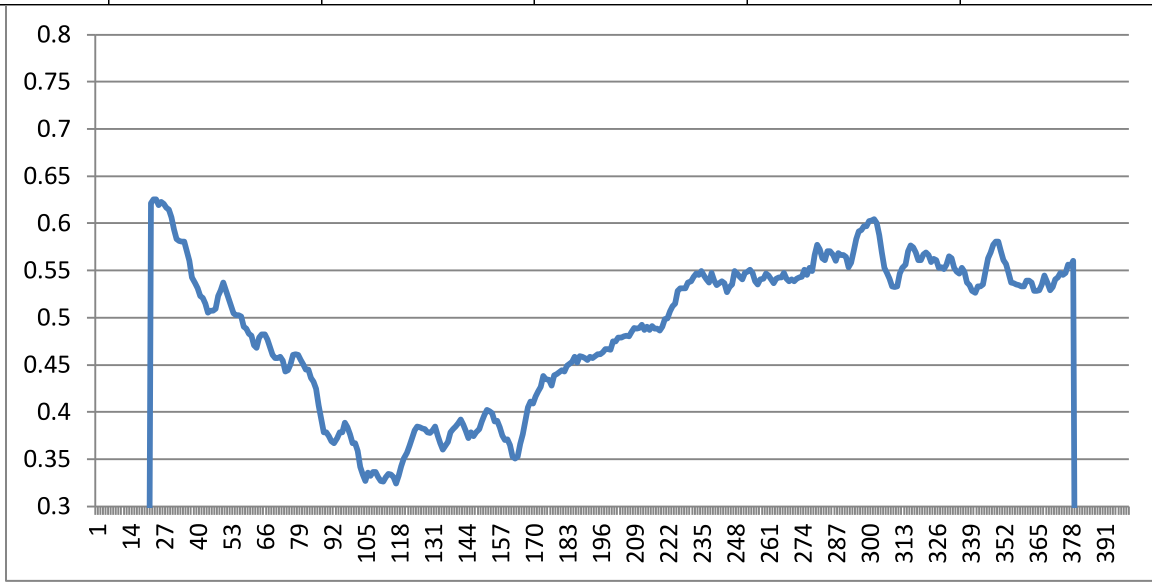


Figure 58 Attentiveness Graph (no baselines used) for Participant 41

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Curious. Waiting for something to happen.	Bored	Confused. Like "that's it?"	Nothing much	Nothing much	Nothing much
Thinking	What's about to happen?	How long was this going to last?	What were they doing in my class?	I was concentrating hard	Realized how simple this was going to be	Bored

Table 28 Thoughts and Feelings as Reported By Participant 41

No EEG connectivity issues for this participant who watched Video1 and was observed to be on task during training (folding paper, etc.). Participants PostQA comments seem to triangulate with attentiveness graph (thinking about what they were missing while on break from class towards the end of the boring video, and the attentiveness not spiking very high in the training as they "realized how simple this was going to be" in the middle. Might have expected the attentiveness to wane even more towards the end of the training based on the comment regarding boredom at the end of the training.

Participant Number 43

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
143	0	1	1	0	0

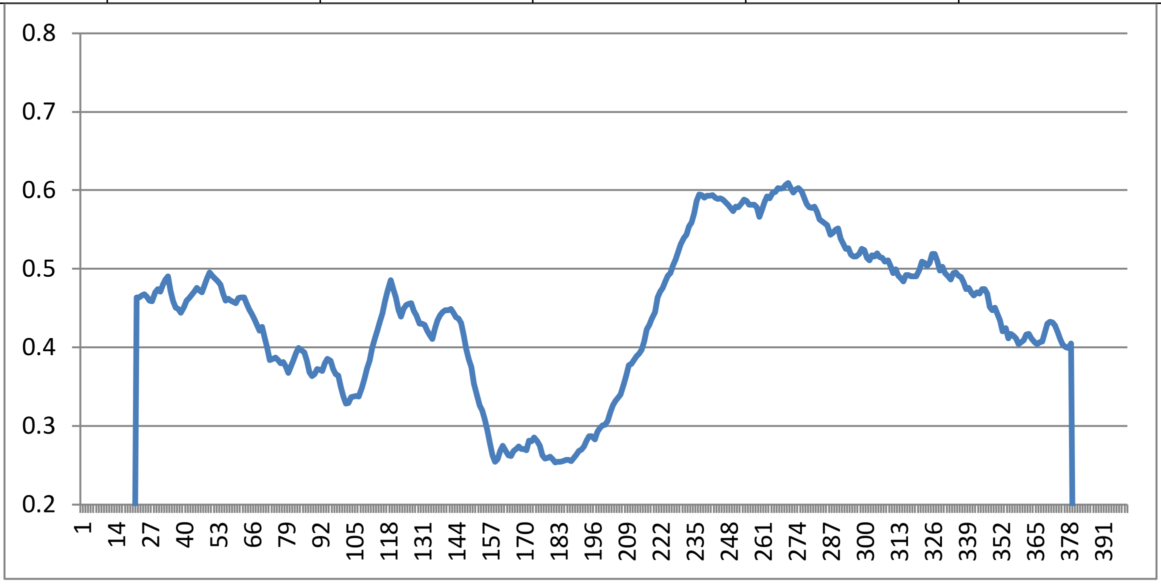


Figure 59 Attentiveness Graph (no baselines used) for Participant 43

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	I felt it was fine and kind of relaxing	I started to feel it was getting bored	I just wanted it to end before I went to sleep watching it	I felt it was going to be interesting	I felt it was going slow pace	I felt instructions were simple elementary school instructions
Thinking	Is it really going to help me relax?	Where was the video taken	How much longer of the video was left	What are we going to do and if I was going to write a lot.	How many times was I going to fold the paper.	Thought it was easy

Table 29 Thoughts and Feelings as Reported By Participant 43

There were no EEG connectivity issues and the participant watched Video1 and was observed to be on task during the training half (folding paper, etc.). The attentiveness graph seems to agree with what would be expected – attentive at first in the boring video, which then wanes quickly. There is a period of attentiveness in the middle of the boring half, and this may be triangulated with the self-described thoughts of the participant at that time “Where was the video taken?" Also, on the training side, the attentiveness wanes rapidly towards the end, and the participant says that they felt the instructions were too simple "elementary school," "easy," "slow pace." This shows the engagement and positive affect is waning for this particular participant.

Attentiveness Graphs examining False Positive Examples

A number of participants were more attentive than the ensemble average during the boring segment of the video. Although most participants were attentive at the very beginning of the boring segment (likely due to the excitement of the experiment beginning) the attentiveness wanes quickly as observed on the attentiveness graph. Below are selections of participants whose attentiveness did not wane, or peaked again during the boring video.

Triangulation with data from the investigators timing notes, as well as participants' thoughts and feelings as described in their own words seem to provide explanations for most, but not all of the false positive situations. The biggest impact on false positives was when there were technical difficulties arising during the boring video. Participants were supposed to see a full screen video of the burning candle (or the boring lecture in the case of Video2a) and otherwise have no distractions. There were technical issues with some participants where the laptop screen had other information on it (recording of EEG data, progress bar and clock showing the position in the video, even one case where the investigator's cellphone vibrated) that can be seen in increased attentiveness measures during those times. Other triangulations between false positives and qualitative data reveal some participants who were extremely nervous, even to the point of verbalizing their fears that the boring video was a trick, and that some frightening apparition would appear (as in the case with some popular internet practical joke videos). There was one situation show below where there is no exceptionally strong triangulation explaining a false positive except that the participant indicated they were thinking about food.

Participant Number 4

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
104	0	0	1	0	0

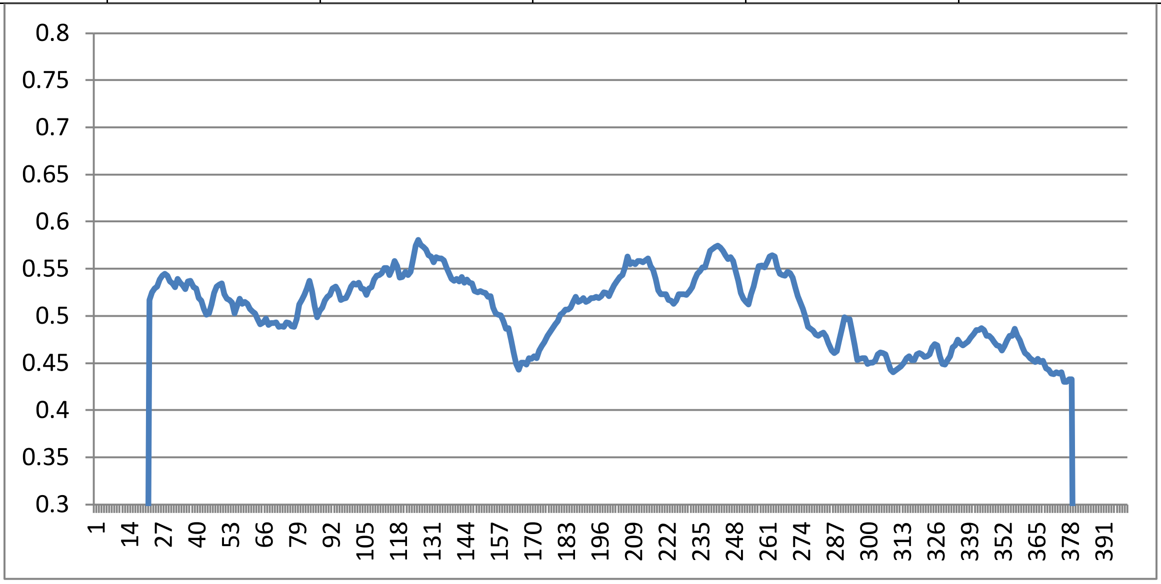


Figure 60 Attentiveness Graph (no baselines used) for Participant 4

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	OK	I was getting more relaxed.	I was feeling sleepy	Comfortable	About the same. I felt more comfortable knowing it was step by step.	Great
Thinking	I first thought was "Is all I'm going to do is watch a candle?"	I was thinking it was starting to get boring.	I could fall asleep.	Looking around	Hoping I got the paper evenly and not cover up word.	I was wondering what was going to be the next step

Table 30 Thoughts and Feelings as Reported By Participant 4

There were no EEG connectivity issues for this participant who watched Video1 and was observed to be on task during the training. Notes from the investigators timing notebook describe the fact that the video was incorrectly not in full screen mode. This allowed the participant to see a portion of the windows operating system info (clock, icons, etc) as well as indicators relating to the collection of the EEG data. The system reported a very high attentiveness for a longer part of the boring video segment than many other participants. This information triangulates well with the participants attentiveness throughout much of the boring video only waning near the end because there were many other things on the laptop screen for the participant to observe and make note of. Also the participant had a clue the training video might be coming up because the video progress bar was visible and nearing the halfway mark (participant had been told that there would be a boring video and then a short training video before the experiment started).

The second note is that the participant did not begin performing the first task in the training video till he turned to the investigator (sitting slightly behind and to the right of the participant) and asked if he should “go ahead” – to which the investigator replied in the affirmative, and from then on the participant was on task. This did not seem to impact the remainder of the attentiveness graph (second half while watching the short training video). It looks similar to the shape of the ensemble average of participants who watched Video1, with peaks near the modeling sections.

Participant Number 13

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
113	0	0	1	0	0

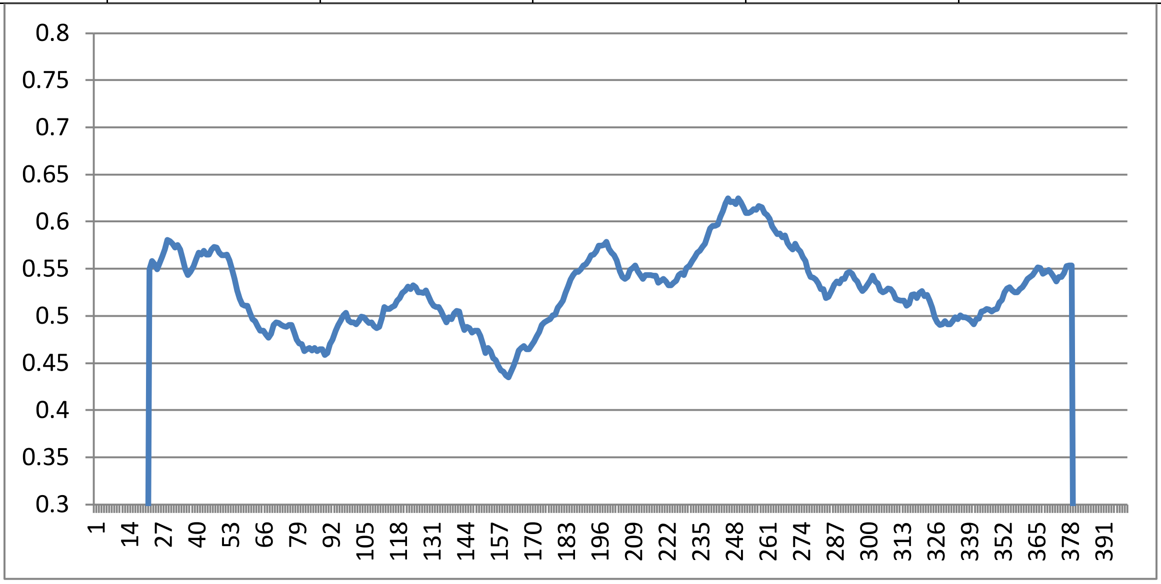


Figure 61 Attentiveness Graph (no baselines used) for Participant 13

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Relaxed and Sleepy	Just trying to get myself mentally ready in any popped up	Words were gone. Something different is going to happen.	It made me feel relaxed with candle, I wanted to get this done.	I don't know. Blank.	Kind of blank. Asking myself why am I doing this? Why couldn't it be another thing to train on?
Thinking	Meditation	Wondering if candle moving. "Nothing's going to pop up is it?"		Ready to see what he would have me do.	Nothing really, just listening	Doing what he asked me to do. It was short.

Table 31 Thoughts and Feelings as Reported By Participant 13

There were EEG connectivity issues which went away after the battery was replaced on the wireless EEG unit. The participant seemed nervous and even during boring half of the video the participant positioned themselves so they could watch the investigator in the reflection of the laptop monitor (the investigator sat behind and to the right of the participant with the timing notebook and stopwatch). In the middle of the boring video the participant turned to the investigator and asked “Nothing’s going to pop up is it?” to which the investigator replied “No.” These comments, as well as the comments in the PostQA by the participant triangulate well with the high attentiveness in the beginning, middle, and end of the boring video (the last area where the participant says they were feeling like “something different was going to happen.” The rest of the attentiveness curve (second half of the video) is much more similar to the ensemble average of the participants who viewed Video1.

Participant Number 38

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
138	0	1	1	0	0

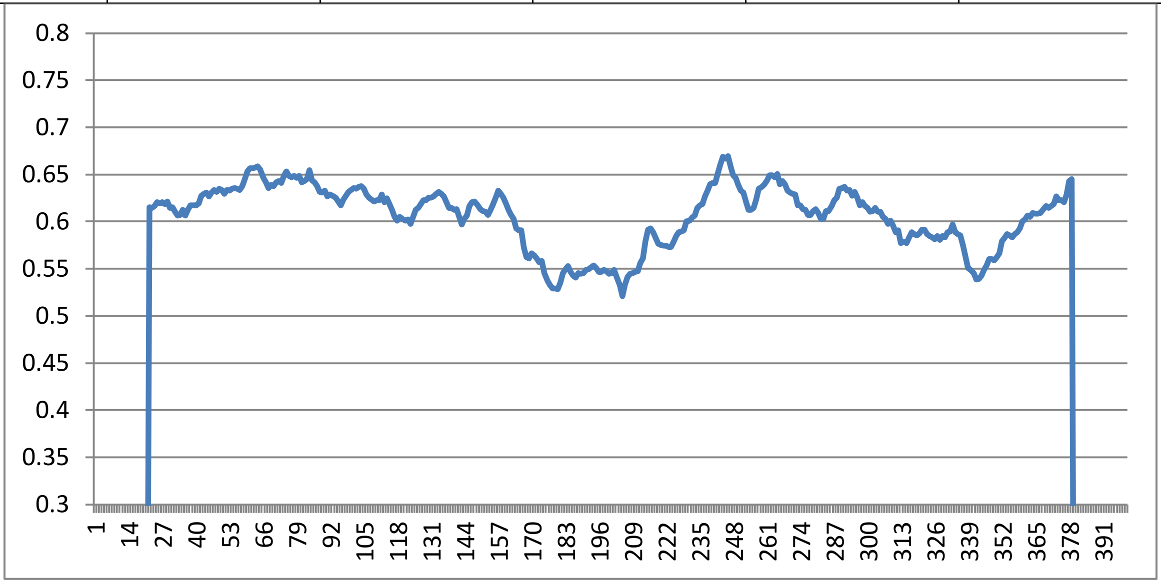


Figure 62 Attentiveness Graph (no baselines used) for Participant 38

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	A little nervous	Bored	Relieved	Simple	Organized	Smart
Thinking	If it was going to be hard	How long is this going to take	Moving on to the next thing	Going back to class	If it was going to get harder	If there was a little more

Table 32 Thoughts and Feelings as Reported By Participant 38

No EEG connectivity issues for this volunteer who watched Video1 and was observed to be on task (folding paper, etc.) during training. The participant was sniffling heavily throughout the entire period of time. Attentiveness was remarkably high throughout the boring half of the video until the end – could be because the participant was self-conscious about sniffling. The participant may have been attentive to the fact that they had a runny nose. The training portion of the video has a more common attentiveness graph, more similar to the ensemble average of all the participants who watched Video1.

Participant Number 56

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
156	0	1	2	0	1

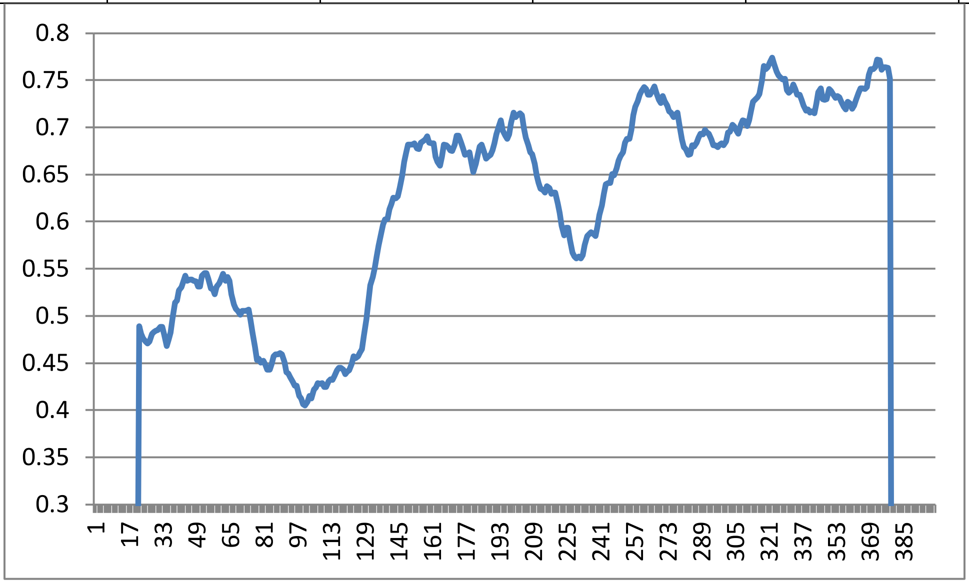


Figure 63 Attentiveness Graph (no baselines used) for Participant 56

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Bored	Bored	Bored	Normal	Normal	Normal
Thinking	How long it will take	Food	Food	What's the purpose	This is boring	Food

Table 33 Thoughts and Feelings as Reported By Participant 56

There were some EEG headset connectivity problems at first but then OK. In this Video2a, the participant watches a boring lecture, and then an attentiveness inducing lecture where the participant was observed to be on task folding paper and such. The attentiveness graph looks pretty normal on the right hand side, but the boring lecture has the end portion measuring relatively high with regard to attentiveness. If baselines were used, the graph would be shifted downwards. There is not enough info to determine why the attentiveness (without baselines) is higher than other participants, but it shares some commonality with the previous participant who also had particularly clean looking EEG data. Perhaps also the participant was interested in the lecture topic, although the PostQA indicated they were bored and thinking about food.

Participant Number 51

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
151	0	1	2	0	1

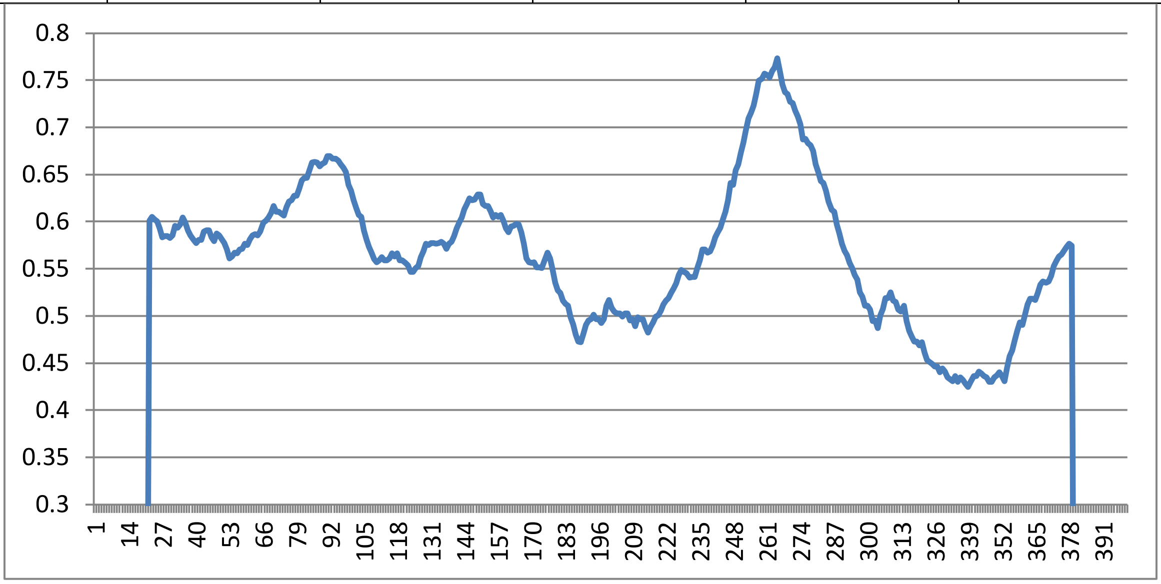


Figure 64 Attentiveness Graph (no baselines used) for Participant 51

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Felt thoughtful and ready to listen	Had wandering thoughts and fell in and out of focus	Bored, lost, unattentive	Attentive	More awake and alert	Satisfied, relieved.
Thinking	What the video could be about	What the main idea of the video was	How did I not retain the video and if I was asked a question I would not be able to answer.	How the video had more color. The bright red shirt.	What the training is for	How easy it was to follow the instructions

Table 34 Thoughts and Feelings as Reported By Participant 51

Participant Number 52

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
152	0	1	3	0	1

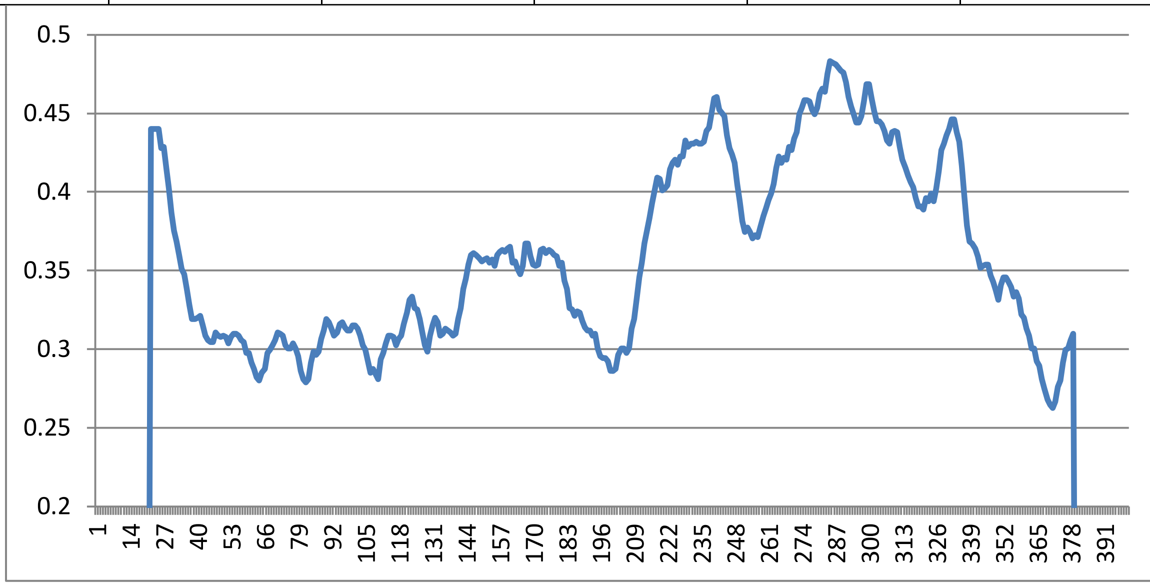


Figure 65 Attentiveness Graph (no baselines used) for Participant 52

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Wondering what was going to take place in the video	Wondered how long would have to look at the candle	Bored	Optimistic, Alert	Impatient, happy	Good inside
Thinking	Looking at color of candle, the light, background	About taking a nap. Lost my interest.	Glad video was over	What the video was going to be about. The number 2 shows	Waiting for the second number. Looking at colors of shirt. Color number card. Looking for surprises.	Waiting for third number. Thinking about Hebrew Algebra

Table 35 Thoughts and Feelings as Reported By Participant 52

No EEG connectivity issues. Participant seemed nervous before the test, asking lots of questions, and telling the investigator that they had participated in many research projects as a subject under Veteran's Affairs. The participant watched Video2b with the candle first and the training video remembering numbers second. The participant was observed to be mostly sitting still throughout (scratched their side 3 minutes into the video), and did remember all three numbers after the video. The baselines would have shifted the attentiveness signal upwards and yielded a very expected attentiveness graph.

Participant Number 55

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
155	0	1	3	0	1

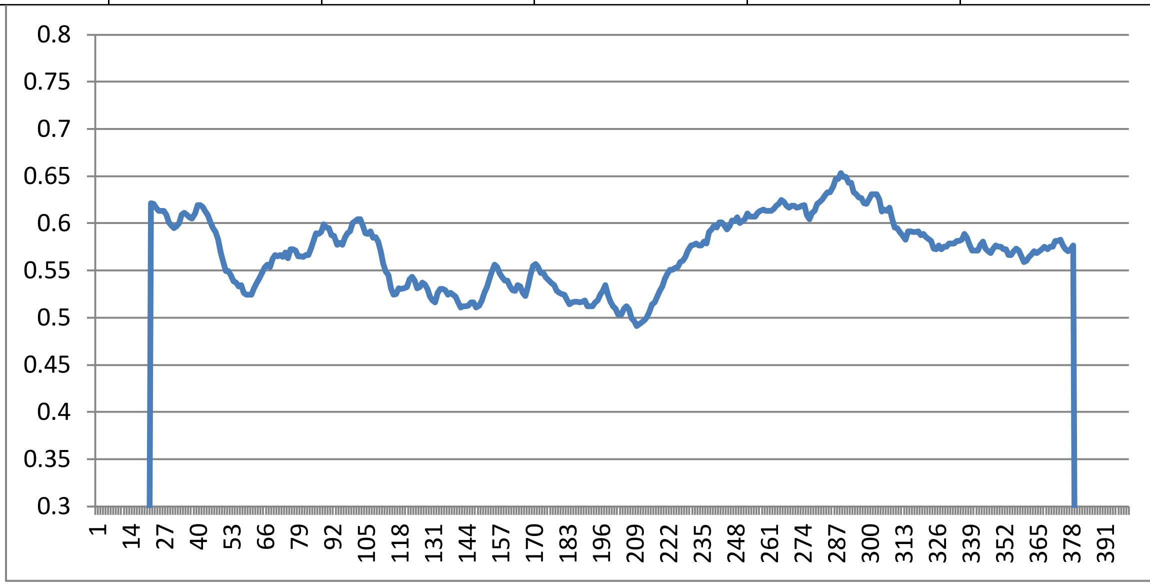


Figure 66 Attentiveness Graph (no baselines used) for Participant 55

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Relaxed	Relaxed	Bored	Nothing	Interested	Nothing
Thinking	Breathing	How long was this going to be?	Where's the interesting part?	What it was about	The numbers	The numbers

Table 36 Thoughts and Feelings as Reported By Participant 55

There were no issues with the EEG data collection, and in this Video2b the participant starts off with the candle boring video and then must remember 3 numbers in the training video. The participant was observed to be sitting still and did remember all three numbers after the video. The attentiveness measures are high throughout both the boring and attentiveness halves, although if baselines were to be used, the graph would have been shifted downwards and would have revealed a rather common attentiveness indicator pattern. There is not enough information to tell why this one participant would have needed the downward shifting of the data. The only indicator might be that the EEG data was noted to have been particularly devoid of noise events (manual visual inspection).

Attentiveness Graphs examining some Noisy EEG Participants

Below are examples of participants that were removed from the training set of data because their EEG signals were manually classified as being excessively noisy. A large frequency of noise spike events was observed, or a regular repetition of noise data was observed that did not seem to be consistent with normal blink or muscular noise data. This qualitative analysis is performed to see if an EEG expert were not available to remove these data sets – what the output of the system would be. As can be seen from the individual results below, future research might be needed to automatically identify excessively noisy EEG data for automatic removal, thereby removing the need for this manual noise removal step.

Participant Number 10

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
110	0	0	1	1	0

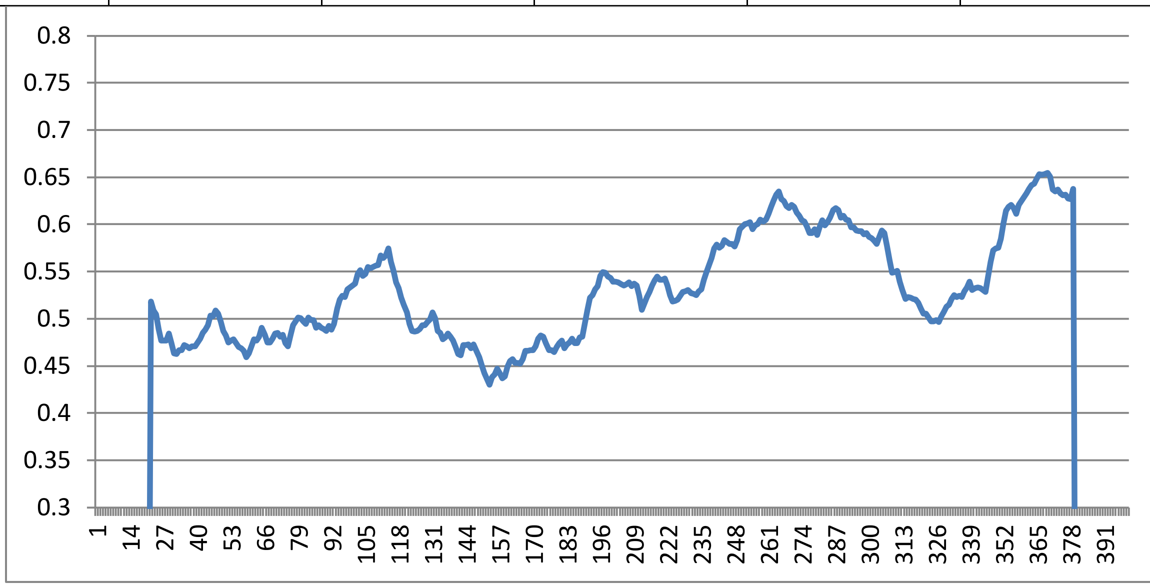


Figure 67 Attentiveness Graph (no baselines used) for Participant 10

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Anxious to see what would happen next; same; change; go away.	Questioned might pop up. Started to analyze different things.	Realized nothing was happen	Excited to see I was actually going to get to interact and do something.	Involved	I felt a sense of accomplishment.
Thinking	Intro something different	Interested to see what's coming next.	How long am I going to sit here?	What was I going to have to do. Tell me then I was going to do something on my own.	That I was going to repeat something	Maybe something else or another video with more steps

Table 37 Thoughts and Feelings as Reported By Participant 10

There were no EEG connectivity issues for this participant who watched Video1 and was observed to be on task (folding paper, etc.) during the training half of the video. The attentiveness curve is very similar to the ensemble average of all the Video1 watching participants, with two notable exceptions. First, there is an attentiveness peak in the middle of the boring video. This triangulates well with the uncommon comment by the participant that their feelings and thoughts in the middle of the boring video included what “might pop up,” and that they were “interested to see what was coming next.” The second deviation from the ensemble average is the strong attentiveness rise at the end of the training video, where other participants’ attentiveness wanes. Here, there is good triangulation with the participants comments about their feelings and thoughts nearing the end of the training where they said that they had a strong positive affect in “a sense of accomplishment” and a focus on the video thinking “maybe something else or another video with more steps.” An interesting point is that this participant was removed from the training set because the EEG signal seemed visually noisy, but nevertheless it had a meaningful attentiveness pattern generated by the DFT system which used the threshold based noise removal algorithm.

Participant Number 12

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
112	0	0	1	1	0

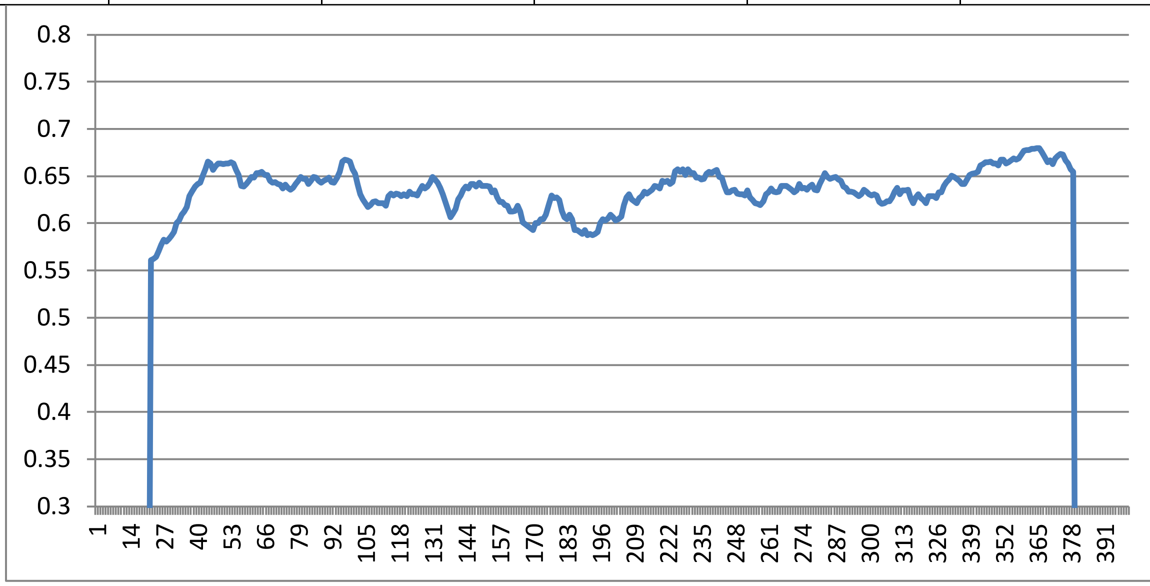


Figure 68 Attentiveness Graph (no baselines used) for Participant 12

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Felt fine. Relaxed	Same thing.	Boring	Fine	Pretty Good, wanted to do it right	I accomplished that
Thinking	I focused on my breathing			Instructions		

Table 38 Thoughts and Feelings as Reported By Participant 12

EEG connectivity issues did not go away even after the batteries changed in the unit. Even when testing the EEG before the beginning of EEG recording (asking the user to blink three times to ensure the EEG is registering correctly) the number of noise spikes was very high, seeming to be at heart rate frequencies. It was as if the EEG sensor was acting like an ECG. Decided to go ahead and continue with the data collection, but error appeared on the screen in front of the video saying “Connection Error” and the investigator had to reach over and operate the laptop keyboard twice to clear the errors (during the training second half of the video). Even though the participant was observed to be on task, the above served as large distractions, and the participant also positioned themselves to be able to see the investigator in the reflection of the laptop (the investigator sat behind and slightly to the left of the participant to take timing notes). All of these triangulate well with the steady level of attentiveness regardless of what was going on in the video, and with the high level of noise in the EEG signal (so noisy that it was discarded as part of the training set).

Participant Number 54

Participant	Self-ID?	New PreQA?	Video Used	Noisy EEG?	Clean EEG?
154	0	1	1	1	0

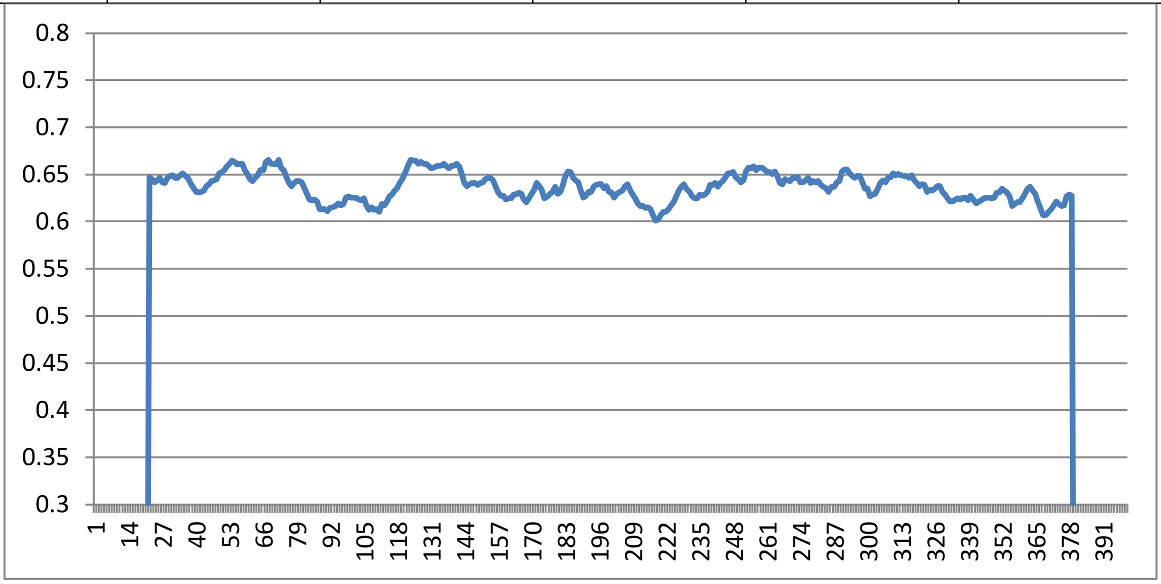


Figure 69 Attentiveness Graph (no baselines used) for Participant 54

	Bor. Begin	Bor. Mid	Bor. End	Tra. Begin	Tra. Mid	Tra. End
Feeling	Relaxed	Relaxing turning to boredom	Somewhat bored	Slight nervousness because I wasn’t sure how many steps it would be.	Some relief because instructions were simple but a little embarrassed when I realized a small mistake	Felt pretty cool I participated in a study.
Thinking	Vacation or going to a spa	Vacations cost money	Almost expecting screen to change as a surprise / jumping actor to OK is this part about done.	Concentrating to make sure I didn’t mess up	Whoops, messed up just a bit because the video said to listen first then do.	Good, quick, easy, and painless.

Table 39 Thoughts and Feelings as Reported By Participant 54

There were no connectivity issues, and the participant watched Video1. The participant wa observed to be on task with regard to folding the paper, etc. The PostQA responses don't seem to indicate anything unusual, however there is very little difference between the boring half and the training half – this can only be explained by the fact that the EEG was extremely noisy (by manual visual inspection) and was deemed to be not useful for training. It is only examined here for triangulation purposes.